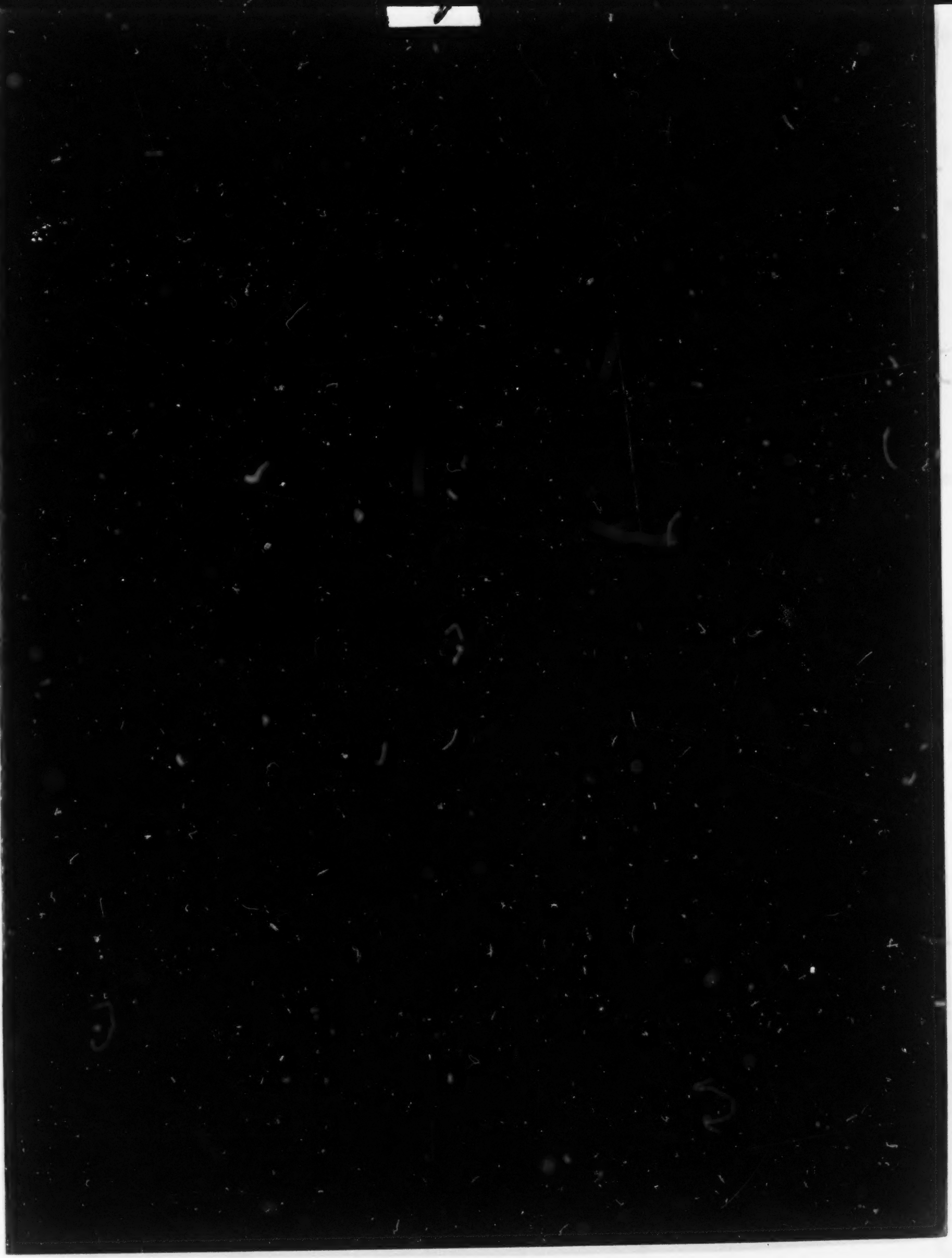
DOCUMENT RESUME

ED 051 823 LI 002 882

TITLE	Urbandoc: A Bibliographic Information System. Technical Supplement 2: Operations Manual.
INSTITUTION	City Univ. of New York, N.Y. Graduate Div.
SPONS AGENCY	Department of Housing and Urban Development, Washington, D.C.
REPORT NO	Urbandoc-71-3
PUB DATE	71
NOTE	250p.
AVAILABLE FROM	Dr. Vivian S. Sessions, Director, Center for the Advancement of Library-Information Science, City University of New York, 9 East 40th Street, New York, N.Y. 10036 (no charge)
EDRS PRICE	EDRS Price MF-$0.65 HC-$9.87
DESCRIPTORS	Computers, Information Processing, *Information Retrieval, *Information Systems, *Operations Research, *Programing, Social Sciences, *Systems Analysis, Urban Renewal
IDENTIFIERS	*URBANDOC Demonstration Project

ABSTRACT

The "Operations Manual" is the third volume of the final report of the URBANDOC Demonstration Project. All three volumes are concerned with the applicability of computerized documentation techniques to the literature of urban planning and renewal. Volume one is the "Demonstration Report" (see LI 002 880), which discusses the goals and accomplishments of the Project, and volume two is the "General Manual" (see LI 002 881), which provides detailed information of the techniques developed for handling the documents. The "Operations Manual" contains detailed systems analysis, programming and operations data information. (Author/MM)



ED051823

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Urbandoc-71-3

Urbandoc/A Bibliographic Information System Technical Supplement 2/Operations Manual

Vivian S. Sessions, Project Director / Lynda W. Sloan, Systems Analyst

HI 002 882

The Urban Renewal Demonstration Project and the publication of this report were made possible through an Urban Renewal Demonstration Grant awarded by the Department of Housing and Urban Development, under the provisions of Section 314 of the Housing Act of 1954, as amended, to The City University of New York.

The Graduate Division, The City University of New York, 1971

AUTHOR ACKNOWLEDGMENTS

Former Staff

The authors gratefully acknowledge their indebtedness to those former members of the URBANDOC staff who were substantially involved in the development of Project URBANDOC:

Document Analysis: Rose Bratton, Patricia Culkin, Rochelle Field, Clifford Johnson, Edith Ward. *Systems and Programming*: Rita Merlin, Robert Najman, Seena Sperling. *Office Assistance*: Mollie Fasbender, Norma Harris, Catherine Laubach, Regina Lewis. *Keypunching*: Nancy Santiago, Elizabeth Williams.

The authors are particularly grateful to Mesdames Bratton, Field and Sperling for their contributions to earlier drafts of the final report, and to Mrs. Harris for preparing the entire final manuscript for submission to the Department of Housing and Urban Development.

Computing Facilities

For assistance in arranging the necessary computing facilities, the project is greatly indebted to Mr. Edwin Brenman of the New York City Bureau of the Budget and to Professor Herbert Arkin of The Barnard Mr. Baruch College of The City University.

University Graduate Division

This project would have been impossible without the whole-hearted support of Dr. Mina Rees, President of the University Graduate Division. Dr. Rees, the Deans, the Business Manager, and members of the administrative staff all contributed to the success of URBANDOC in so many ways that the authors find themselves overwhelmed in trying to adequately express their thanks.

Research Foundation

The authors are also indebted to The Research Foundation of The City University for its role in managing Project URBANDOC.

Report Design

Dean Marilyn Mikulsky, The City University Graduate Division

PREFACE

Project URBANDOC is reporting on four years of activity as an Urban Renewal Demonstration Project at The City University of New York. The project evolved from a need for improving bibliographic services in urban affairs – and specifically urban renewal – at a time when computer technology was being incorporated into a wide range of information systems. URBANDOC was one of the first of the library-information science systems to deal specifically with the social sciences.

The final report consists of three volumes: the *Demonstration Report,* the *General Manual* (Technical Supplement 1), and the *Operations Manual* (Technical Supplement 2). Each of these is bound separately and intended for separate distribution. For the most general reader who wishes an over-all view of the objectives, features, accomplishments, and conclusions and recommendations of the project, the *Demonstration Report* should suffice.

The *General Manual* is designed to provide the reader with detailed knowledge of the techniques developed for handling the documents according to library-information science practices as developed by Project URBANDOC. While it also provides an overview of the programming system used by the project, the *Operations Manual* should be consulted for detailed systems analysis, programming, and operations data.

The U.S. Department of Housing and Urban Development has been most generous in its assistance of Project URBANDOC, from project submission to final report. HUD's committment to the Demonstration was as important conceptually as it was economically, and the University's indebtedness is thus two-fold. The President and Deans of the University Graduate Division join the New York City Planning Commission and the URBANDOC staff in thanking the Department for having made possible each of these three final volumes, as well as the entire project.

TABLE OF CONTENTS

Introduction / Chapter I 1- 4

Systems Background 1
Computer Equipment 2
Organization of the Manual 3

Thesaurus Module / Chapter II 5-18

The Thesaurus File 5
Program Inventory 7
Program Abstracts 10
Input Specifications 14
Tape File Specifications 15
References to the URBANDOC Final Report 17

Pre-edit Module / Chapter III 19-29

Program Inventory 19
Program Abstracts 20
Input Specifications 22
Tape File Specifications 25
References to the URBANDOC Final Report 29

File Maintenance Module / Chapter IV 30-54

The Document Master File 30
The Inverted File 33
Program Inventory 36
Program Abstracts 38
Tape File Specifications 45
References to the URBANDOC Final Report 53

Search Module / Chapter V 55-83

Introduction 55
Narrative 59
Program Inventory 67
Program Abstracts 69
Input Specifications 73
Tape File Specifications 75
References to the URBANDOC Final Report 82

Publications Module / Chapter VI 84-92

The Publication File 84
Program Inventory 84
Program Abstracts 85
Tape File Specifications 90
References to the URBANDOC Final Report 91

Data Entry / Chapter VII 93-135

Thesaurus Data Entry 93
Thesaurus Worksheet – Subject Descriptors, Subdescriptors 93
Thesaurus Worksheet – Geographic Descriptors 102
Thesaurus Revisions 108
Program Control Card – Thesaurus Worksheets 110

Document Data Entry 110
Document Worksheet 110

Editing and Validation 115
Document Master File Revisions 122
Program Control Card – Document Worksheet 127

Search Data Entry 127
Basic Search Worksheet. 127
Expanded Search Worksheet 130
Program Control Card – Search Worksheets 134

Processing Cycles / Chapter VIII 136-148
Editing and Validation Cycle 136
Input Processing Cycle 137
Input Index File Maintenance Cycle 140
Input Index Production Cycle 140
Retrieval Report File Maintenance Cycle. 143
Retrieval Report Production Cycle 143
Miscellaneous Thesaurus Products 146
Miscellaneous Document Products 146
Miscellaneous Publications Products 147

Operating Instructions / Chapter IX 149-215
Thesaurus Module 149
Pre-edit Module 157
File Maintenance Module. 162
Search Module 180
Publications Module 194
SORT7 Halt List 208

Error Listings and Systems Message / Chapter X 216-228
Thesaurus Module 216
Pre-edit Module 218
File Maintenance Module. 220
Search Module 224
Publications Module 228

Tape Library and Report Controls, Restart Procedures / Chapter XI 229-234
Thesaurus Module 229
Pre-edit Module 230
File Maintenance Module 231
Search Module 232
Publications Module 233

Timing / Chapter XII 235-241
Thesaurus Module 235
Pre-edit Module 236
File Maintenance Module. 236
Search Module 237
Publications Module 239
Sort Timing Tables 240

Local Implementation / Chapter XIII 242

Bibliography / Chapter XIV 243
Applications Programs 243
Programming Systems 243

TABLE OF FIGURES

Figure 1.	Thesaurus File Record Layout	8
Figure 2.	Machine-Language Codes for Core Storage Addresses	9
Figure 3.	Document Master File Record Layout	34
Figure 4.	Inverted File Record Layout	35
Figure 5.	Concept of the Combined File Search Method	57
Figure 6.	Logical Flow of Combined Search Process	58
Figure 7.	CFS Output and Format Specifications	65
Figure 8.	Publication File Record Layout	86
Figure 9.	Thesaurus Worksheet — Subject Descriptors, Subdescriptors	100-101
Figure 10.	Thesaurus Worksheet — Geographic Descriptors	109
Figure 11.	Document Worksheet	116-118
Figure 12.	Basic Search Worksheet	131
Figure 13.	Expanded Search Worksheet	135
Figure 14.	Editing and Validation Cycle	138
Figure 15.	Input Processing Cycle	139
Figure 16.	INPUT INDEX File Maintenance Cycle	141
Figure 17.	INPUT INDEX Production Cycle	142
Figure 18.	RETRIEVAL REPORT File Maintenance Cycle	144
Figure 19.	RETRIEVAL REPORT Production Cycle	145

INTRODUCTION

Systems Background

The *General Manual* states the objectives of the computer component of the URBANDOC project as:

> To build and maintain a thesaurus in machine-readable form;
>
> To build and maintain a document master file of bibliographic records containing both the content and descriptive analyses in machine-readable form;
>
> To organize the machine-readable files for retrieval and publications purposes;
>
> To establish search programs to query the document master file;
>
> To establish publications programs to produce listings and indexes to these listings.

The system realizing these objectives consists of five subsystems or modules, a module being defined as all the programs required to meet a specific objective. The modules as finally developed are Pre-edit, Thesaurus, File Maintenance, Search and Publications. The programs included in each module came from a variety of sources: URBANDOC programs, the IBM Program Library, and the *Engineering Index.* Each source and the role it played will be discussed in detail.

The core of the system was a set of computer programs from the IBM Program Library, the Combined File Search System, often referred to as CFS. It seemed to fulfill the majority of URBANDOC'S requirements in the areas of thesaurus, file maintenance and search, as well as allow for a gradual expansion to a total systems approach. This set of programs was designed for use on the IBM 1401 computer. (For details of configuration, see section following on Computer Equipment.)

The Pre-edit and Publications Modules were completely developed and implemented by URBANDOC. By adding several programs which were not part of the system as distributed by the IBM Program Library, URBANDOC enlarged the nature of the Thesaurus Module and File Maintenance Module. The Thesaurus Module also included another program from the IBM Program Library, the one which permutes the Thesaurus. The Search Module was expanded by the incorporation of the search subsystem acquired from the *Engineering Index*. Throughout the entire system, sorting of the data files is performed by SORT7, another program from the IBM Program Library.

The source of a program has a direct bearing on the issues of support and maintenance. While URBANDOC was responsible for supporting the programs in its system, the degree of effort expended was directly related to the source of the program. URBANDOC was best able to maintain the programs it had created. The maintenance of the programs from the *Engineering Index* and the IBM Program Library involved other factors, namely that

as users rather than primary developers, the project was less familiar with the detailed workings of each program and, in some instances, less able to effectuate all changes.

The programs from the IBM Program Library were classified as "Type III programs". These were programs donated to the library by their corporate developers which were distributed on request to users of IBM equipment. IBM had no maintenance committment for these systems. The documentation accompanying the programs usually included the names of individual authors who could be contacted in case of problems or questions. Of course, such programs as SORT7, the COBOL compiler, and the AUTO-CODER assembler were Type I programs which were completely supported by IBM.

The task of program maintenance for permuting the Thesaurus was an issue that never arose; URBANDOC did not encounter operating difficulties with that particular program. The situation in the case of the CFS programs was entirely different. URBANDOC resolved many of the issues by itself and a CFS Users' Group was formed to cooperatively solve the remaining issues. (It is interesting to note that approximately seven hundred changes were made to the CFS system with an estimated average of two man-hours required per change.)

With regard to the *Engineering Index* programs, some minor format changes were made to produce retrieval output in agreement with URBANDOC terms rather than the ones used by the *Engineering Index*. In other instances, the *Engineering Index* did provide URBANDOC with assistance in learning to use the search subsystem to its best advantage. In this respect, the assistance was more a matter of training rather than resolving programming problems.

With regard to prospective support and maintenance of the various programs, none of the previous arrangements can be relied upon for assistance. *Engineering Index* is no longer actively supporting these programs; in some instances it may be possible to make separate arrangements with them for additional support and/or maintenance. The persons originally named on programs in the IBM Program Library have long since lost intimate familiarity with these particular programs and have perhaps moved to other projects, organizations, etc. URBANDOC cannot provide future support for the system unless special arrangements are made with The City University Graduate Division.

Computer Equipment

URBANDOC has been operating its system on the IBM 1401 computer at the Baruch College of The City University of New York. URBANDOC does not use all the features and devices in this particular configuration. The portions that it does use are the following:

12,000 positions of computer memory (12K);
a model 1402 card reader-punch;
a model 1403 printer;
four tape drives;
one model 1311 disk drive with diskpacks;
advanced programming special feature;

high-low-equal compare special feature; and
sense switches.

Except for the Publications Module, one program in the Pre-edit Module and the *Engineering Index* programs, the system could be operated on a machine with the following features and devices:

8,000 positions of computer memory (8K);
a model 1402 card reader-punch;
a model 1403 printer;
four tape drives;
advanced programming special feature;
high-low-equal compare special feature; and
sense switches

(See the program inventory in each module discussion for the specific configuration applicable to each program.)

As noted above, any model tape drive on the 1401 can be used without performing systems changes provided that the tapes are created at a density of 556 characters per inch (cpi). If a density of 800 cpi is used, the sort control cards must be changed to reflect the new tape density. (See Chapter IX, Operating Instructions.)

URBANDOC uses the 1311 disk drive to facilitate searching (see Chapter V, specifically the abstract for PHASE4, and Chapter IX, specifically the operating instructions for PHASE4) and to assemble or compile new programs under development. Searching can be performed without the use of a disk drive and programs can be assembled using the tape version of COBOL and AUTOCODER translators.

Other special features such as multiply-divide, print storage and a model 1407 inquiry console might exist on other systems. The multiply-divide feature is not necessary to the operation of the system. If present, it produces no added advantages. The same holds true for the inquiry console. However, the presence of the print storage feature could serve to reduce run time for programs with heavy printing.

URBANDOC's system could also be run on an IBM 1460, an IBM 1410 with 1401 compatibility or an IBM 360 with the 1401 compatibility feature. It might also be possible to operate the system on other manufacturers' equipment (ie., the Honeywell 200) through the use of a translator like "Liberator". However, URBANDOC has not tested any of the above procedures and could not guarantee their success.) The above statements on machine-independence assume that other systems meet the input-output requirements concerning a card reader-punch, printer, and tape drives.

Organization of the Manual

The *Operations Manual* is organized according to usage: Chapter II through Chapter VI for systems analysis and programming, Chapter VII through Chapter XIII for operations. Systems analysis and programming are, in turn, organized by program module. A program

inventory and detailed specifications of the input files and both major and auxiliary data files are described in the overview of each program module. Abstracts and programmers' notes (where applicable) are also provided for each program for a complete discussion of each module.

Operations are divided into several areas. Data Entry, Processing Cycles, Operating Instructions and Error Listings are part of the daily operation of the system. Timing and Tape Library and Report Controls are part of the management of the system.

THESAURUS MODULE

The Thesaurus File

The Thesaurus File is the control file for the content analysis terms, that is terms from the URBANDOC Thesaurus that are assigned by the document analyst to describe a document's contents. (See the *General Manual,* Chapter IV, specifically the Subject Thesaurus.) The Thesaurus File is used primarily to validate the content analysis terms, to indicate the correct usage of a term to the document analyst and to make the substitution from the natural language geographic term to code form. (See the *General Manual,* Chapter IV, specifically the Geographical Thesaurus.) The Thesaurus File itself contains two types of terms: subject and geographic. From a data processing viewpoint, no distinction is made between the two types.

CFS systems requirements state that all the information for one term be contained within a single tape record. URBANDOC has developed a technique for the extension of term information to several tape records. (This will be discussed under continuation records.) In all future discussions, however, a reference to a term is to one tape record. The system treats continuation records as "new terms".

When considering type of information on the Thesaurus File, the reference is to the sections (Term, Cross-reference and Program Utility) within a single tape record. The computer system uses the Term Section to perform the computer validation of terms and the substitution of the geographic codes. The Cross-reference Section contains the data indicating to the document analyst the correct usage of the terms; this information is not processed but only printed by the computer system. The Program Utility Section is included to facilitate computer processing during file maintenance. This information is used only by the programs and is not made available either to the document analyst or external user of the Thesaurus.

With no distinction made between the subject and geographic information, there is only one type of tape record on the file. The maximum size of this tape record is 1296 characters. The Program Utility Section and Term Section are fixed in structure and size. Only the Cross-reference Section is variable in length.

The Program Utility Section is contained in every record. The first section on the tape record, it contains the following information about the size and nature of the term being entered:

Record size: a three-position field containing the number of characters in the record. For records with more than 999 characters of information, the field is maintained in 1401 machine-language code for core storage addresses. (See Figure 2.);

Term size: a three-position field containing the number of characters in the actual term field. A term may be variable in length to a maximum of fifty-four characters;

Cross-reference size: a three-position field containing the number of characters of

cross-reference data. For records with more than 999 characters of cross-reference information, the field is maintained in 1401 machine-language code for core storage addresses;

Substitution flag: a one-position code indicating the presence of substitute information for a term;

Truncation flag: a one-position flag indicating the presence of a term root;

Entry code: a one-position code for the authorized usage of a term as a descriptor, subdescriptor or both.

The Term Section is composed of the authorized term, the preferred form (or substitute) for geographic terms and synonyms, the date of the term's entry into the Thesaurus and its last date of revision. (For the formation of these fields, see Chapter VII, specifically Thesaurus Data Entry.) It also includes type code indicating the authorized usage of a descriptor. This section immediately follows the Program Utility Section in the record.

Type code indicates whether a descriptor may be used for independent searching or if it must be used in combination with another term. Since subdescriptors cannot be searched independently, this field is blank if the term is a subdescriptor. It is always used for a descriptor. (For the contents of this field, see Chapter VII, specifically Thesaurus Data Entry.)

The substitute entries for a term (if used) include a search substitute and a publications substitute. The search substitute is used to replace an input term used as a descriptor or subdescriptor. The publications substitute is used to replace an input term used as a subject heading. The original CFS system made provision for just a search substitute. Since this caused problems when the term was used as a subject heading, the publications substitute was created which eliminated all coding and enlarged the size of the substituted field. These fields will be used only if the substitution flag has been set. (For further information on the format of each substitute, see Chapter VII, specifically Thesaurus Data Entry.)

The *General Manual* contains a detailed discussion of the formation of the descriptor, subdescriptor and publications substitute. Each term will be stored left-justified in the field and the remainder of the field right-padded with blanks if necessary. The fields are stored on the record exactly as entered onto the worksheet. (For further information, see Chapter VII, specifically Thesaurus Data Entry. See also the *General Manual*, Chapter VI, specifically the Thesaurus File Description.)

The Cross-reference Section is not processed but only printed by the computer system. This is information that may be entered by the document analyst for use by himself or an external user. The contents and format are unstructured except that within one record this section cannot exceed sixteen lines of information and each line may not exceed seventy-two characters in length. The system does not actually count the number of characters included on a card until it reaches the maximum limit. Rather, it accepts the first sixteen cards and stores seventy-two characters of information from each.

The Cross-reference Section may contain such information as the relationship of a term to other terms in the Thesaurus or scope notes. Within each line of information, the data is stored exactly as entered onto the worksheet. This data is not edited other than for length limitations. The Cross-reference Section immediately follows the Term Section in the record. (For further information as to the contents of this section, refer to Chapter VII, specifically Thesaurus Data Entry.)

In certain instances, there are more than sixteen cross-references for a term. In this case, a continuation record must be created for each additional sixteen cross-references. The continuation record will not be processed but only printed by the system. The basic term record will contain the term and the first sixteen cross-references. The continuation record will contain a continuation note as part of the term since more than one continuation record may be necessary, for example: FORMATS (CONTA) and the next sixteen cross-references.

Program Inventory

Program Number	Program Function	Source	Basic or On-demand*	Programming Language	Configuration Set**
XMAIN0	To provide a pre-listing of the Thesaurus Input	URBANDOC	Basic	Disk AUTOCODER	Either 1 or 2
XMAIN1	Tc format the Thesaurus Input	CFS	Basic	Tape AUTOCODER	Either 1 or 2
XMAIN2	To update the Thesaurus File	CFS	Basic	Tape AUTOCODER	Either 1 or 2
XMAIN3	To print the Thesaurus File, subject section or geographic section	CFS	Basic	Tape	Either 1 or 2
XMAIN4	To produce the Statistical Analysis of the Thesaurus File	URBANDOC	Basic	Tape AUTOCODER	Either 1 or 2
XMAIN5	To print the Permuted Thesaurus File	IBM Program Library	On-demand	Tape AUTOCODER	Either 1 or 2
XMAIN6	To create a Thesaurus Supplement File	URBANDOC	On-demand	Disk AUTOCODER	Either 1 or 2

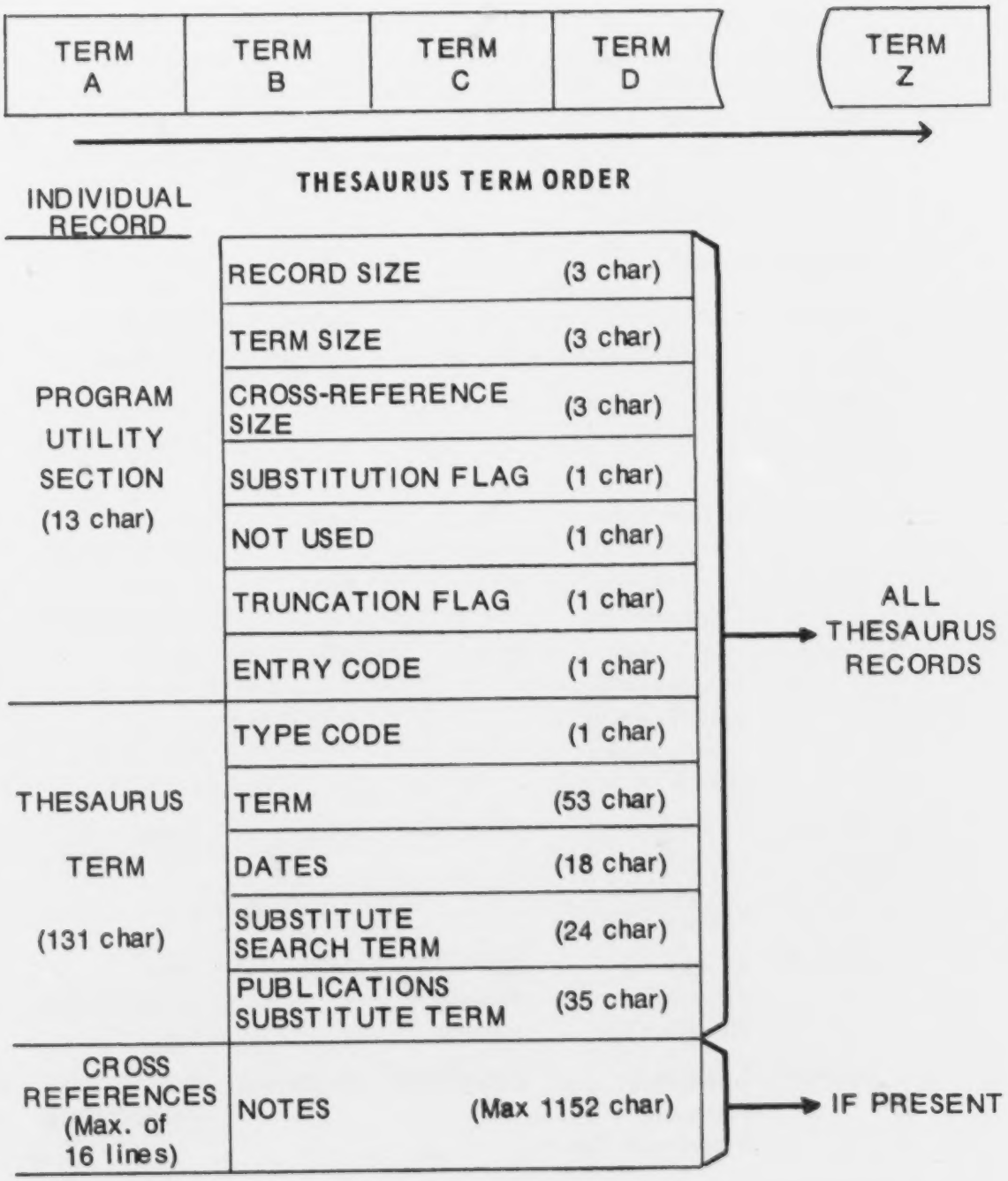
*Basic: Part of the standard processing for updating the Thesaurus File.

On-demand: Processing performed on request only.

**Set 1: 12K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; 1311 disk drive; advanced programming; high-low-equal compare; sense switches.

Set 2: 8K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; advanced programming; high-low-equal compare; sense switches.

THESAURUS FILE RECORD LAYOUT[1]



[1] Reproduced from Chapter VI of the *General Manual*.

Figure 1

MACHINE–LANGUAGE CODES FOR CORE STORAGE ADDRESSES[2]

	ADDRESSES 0000-3999		ADDRESSES 4000-7999 A-Bit (0-Zone) over Units Position		ADDRESSES 8000-11999 B-Bit (11-Zone) over Units Position		ADDRESSES 12000-15999 AB-Bits (12-Zone) over Units Position	
	Addresses	Codes	Addresses	Codes	Addresses	Codes	Addresses	Codes
	0000-0099	000-099	4000-4099	00 ǂ -09Z	8000-8099	00!-09R	12000-12099	00?-09I
	0100-0199	100-199	4100-4199	10 ǂ -19Z	8100-8199	10!-19R	12100-12199	10?-19I
	0200-0299	200-299	4200-4299	20 ǂ -29Z	8200-8299	20!-29R	12200-12299	20?-29I
	0300-0399	300-399	4300-4399	30 ǂ -39Z	8300-8399	30!-39R	12300-12399	30?-39I
	0400-0499	400-499	4400-4499	40 ǂ -49Z	8400-8499	40!-49R	12400-12499	40?-49I
	0500-0599	500-599	4500-4599	50 ǂ -59Z	8500-8599	50!-59R	12500-12599	50?-59I
	0600-0699	600-699	4600-4699	60 ǂ -69Z	8600-8699	60!-69R	12600-12699	60?-69I
	0700-0799	700-799	4700-4799	70 ǂ -79Z	8700-8799	70!-79R	12700-12799	70?-79I
	0800-0899	800-899	4800-4899	80 ǂ -89Z	8800-8899	80!-89R	12800-12899	80?-89I
	0900-0999	900-999	4900-4999	90 ǂ -99Z	8900-8999	90!-99R	12900-12999	90?-99I
A-Bit (0-Zone) over Hundreds Position	1000-1099	ǂ 00-ǂ 99	5000-5099	ǂ 0 ǂ -ǂ 9Z	9000-9099	ǂ 0!-ǂ 9R	13000-13099	ǂ 0?-ǂ 9I
	1100-1199	/00-/99	5100-5199	/0 ǂ -/9Z	9100-9199	/0!-/9R	13100-13199	/0?-/9I
	1200-1299	S00-S99	5200-5299	S0 ǂ -S9Z	9200-9299	S0!-S9R	13200-13299	S0?-S9I
	1300-1399	T00-T99	5300-5399	T0 ǂ -T9Z	9300-9399	T0!-T9R	13300-13399	T0?-T9I
	1400-1499	U00-U99	5400-5499	U0 ǂ -U9Z	9400-9499	U0!-U9R	13400-13499	U0?-U9I
	1500-1599	V00-V99	5500-5599	V0 ǂ -V9Z	9500-9599	V0!-V9R	13500-13599	V0?-V9I
	1600-1699	W00-W99	5600-5699	W0 ǂ -W9Z	9600-9699	W0!-W9R	13600-13699	W0?-W9I
	1700-1799	X00-X99	5700-5799	X0 ǂ -X9Z	9700-9799	X0!-X9R	13700-13799	X0?-X9I
	1800-1899	Y00-Y99	5800-5899	Y0 ǂ -Y9Z	9800-9899	Y0!-Y9R	13800-13899	Y0?-Y9I
	1900-1999	Z00-Z99	5900-5999	Z0 ǂ -Z9Z	9900-9999	Z0!-Z9R	13900-13999	Z0?-Z9I
B-Bit (11-Zone) over Hundreds Position	2000-2099	!00-!99	6000-6099	!0 ǂ -!9Z	10000-10099	!0!-!9R	14000-14099	!0?-!9I
	2100-2199	J00-J99	6100-6199	J0 ǂ -J9Z	10100-10199	J0!-J9R	14100-14199	J0?-J9I
	2200-2299	K00-K99	6200-6299	K0 ǂ -K9Z	10200-10299	K0!-K9R	14200-14299	K0?-K9I
	2300-2399	L00-L99	6300-6399	L0 ǂ -L9Z	10300-10399	L0!-L9R	14300-14399	L0?-L9I
	2400-2499	M00-M99	6400-6499	M0 ǂ -M9Z	10400-10499	M0!-M9R	14400-14499	M0?-M9I
	2500-2599	N00-N99	6500-6599	N0 ǂ -N9Z	10500-10599	N0!-N9R	14500-14599	N0?-N9I
	2600-2699	O00-O99	6600-6699	O0 ǂ -O9Z	10600-10699	O0!-O9R	14600-14699	O0?-O9I
	2700-2799	P00-P99	6700-6799	P0 ǂ -P9Z	10700-10799	P0!-P9R	14700-14799	P0?-P9I
	2800-2899	Q00-Q99	6800-6899	Q0 ǂ -Q9Z	10800-10899	Q0!-Q9R	14800-14899	Q0?-Q9I
	2900-2999	R00-R99	6900-6999	R0 ǂ -R9Z	10900-10999	R0!-R9R	14900-14999	R0?-R9I
AB-Bits (12-Zone) over Hundreds Position	3000-3099	?00-?99	7000-7099	?0 ǂ -?9Z	11000-11099	?0!-?9R	15000-15099	?0?-?9I
	3100-3199	A00-A99	7100-7199	A0 ǂ -A9Z	11100-11199	A0!-A9R	15100-15199	A0?-A9I
	3200-3299	B00-B99	7200-7299	B0 ǂ -B9Z	11200-11299	B0!-B9R	15200-15299	B0?-B9I
	3300-3399	C00-C99	7300-7399	C0 ǂ -C9Z	11300-11399	C0!-C9R	15300-15399	C0?-C9I
	3400-3499	D00-D99	7400-7499	D0 ǂ -D9Z	11400-11499	D0!-D9R	15400-15499	D0?-D9I
	3500-3599	E00-E99	7500-7599	E0 ǂ -E9Z	11500-11599	E0!-E9R	15500-15599	E0?-E9I
	3600-3699	F00-F99	7600-7699	F0 ǂ -F9Z	11600-11699	F0!-F9R	15600-15699	F0?-F9I
	3700-3799	G00-G99	7700-7799	G0 ǂ -G9Z	11700-11799	G0!-G9R	15700-15799	G0?-G9I
	3800-3899	H00-H99	7800-7899	H0 ǂ -H9Z	11800-11899	H0!-H9R	15800-15899	H0?-H9I
	3900-3999	I00-I99	7900-7999	I0 ǂ -I9Z	11900-11999	I0!-I9R	15900-15999	I0?-I9I

Units Position: Address Digit	Code	Units Position: Address Digit	Code	Units Position: Address Digit	Code
0	ǂ	0	I	0	?
1	/	1	J	1	A
2	S	2	K	2	B
3	T	3	L	3	C
4	U	4	M	4	D
5	V	5	N	5	E
6	W	6	O	6	F
7	X	7	P	7	G
8	Y	8	Q	8	H
9	Z	9	R	9	I

[2] International Business Machines, Data Processing Division, *IBM 1401 Data Processing System Operator's Guide*, 151p. (White Plains, N.Y., 1965), p140.

Figure 2

Program Abstracts

XMAIN0 — Pre-list of the Thesaurus input

Abstract

The pre-list was designed to detect errors in the Thesaurus Input through computer checking and visual verification prior to the Thesaurus File update. The pre-list provides proof-reading copy along with diagnostics of computer-detected errors.

Programmer's Notes

Definition of a term set. A term set is defined as all the units with the same accession number. (See Chapter VII, specifically Thesaurus Data Entry.)

Use of accession number for changing a term. When changing a term on the Thesaurus File, the transaction is handled by the deletion of the existing term and the entry of the term in its new form. The entry of consecutive term sets with the same accession number cannot be properly edited by the program. Since the deletion unit and the revised term set have the same accession number, it is important to separate the deletion entries from the additions in the input file.

Differentiation between additions and deletions. Even though the additions and deletions are kept separate, they can still be processed as part of the same run. Rather than treating them as two decks of cards, they are to be considered as two portions of one deck. Although the system does not specify which is to appear first, URBANDOC places the deletions before the additions to facilitate SORTX1.

XMAIN1 — Formatting the Thesaurus Input

Abstract

The Thesaurus Input, as first entered, is on cards. XMAIN1 edits the term sets for sequence, coding, etc. (See Chapter VII, specifically Thesaurus Data Entry.) The terms will be compacted and the term sets without errors formatted into single records. The terms which did not meet the editing standards will be included in the Error Listing and removed from further processing.

Programmer's Notes

Differentiation between additions and deletions. When changing a term on the Thesaurus File, the transaction is handled by the deletion of the existing term and the entry of the term in its new form. Because of this, as discussed in XMAIN0, it is important to keep the deletion entries separate from the additions in the input file.

Use of the accession number. Accession number is used only to tie together the various units of a term set in the input. Once the term set has been edited and a tape record created, accession number is dropped. Future processing will use the term itself as the control field for the file.

Compaction of the term. CFS allows the document analyst to enter a term anywhere within the term field on the Thesaurus Worksheet. As part of the formatting procedure, the term will be stored at the left-most boundary of the field (left-justified within the field). In the process of locating and shifting the term, any leading blanks and consecutive embedded blanks (after the first) will be removed. The character content of the term is not examined here. (For further information, see *General Manual,* Chapter VI, specifically the Thesaurus File Description.) The length of the term is not affected other than for the removal of extra blank characters.

SORTX1 – Sort the Thesaurus Input File

Abstract

Since the Thesaurus Input File will update the Thesaurus File, it must be sorted into alphabetical sequence by term. The Thesaurus Input is entered as two blocks of data, deletions and additions. Within each block, the entries appear in random alphabetical sequence by term. At the conclusion of the sort, the file will be in sequence by term (according to the computer collating sequence of the IBM 1401 as illustrated in Figure 3). For a term change in which there is both a deletion entry and an addition entry, the deletion entry will be placed before the addition entry.

XMAIN2 – Create and/or Update the Thesaurus File

Abstract

The Thesaurus File will be updated with the contents of the Sorted Thesaurus Input File. Existing terms may be deleted. New terms may be added. The contents of an existing term set (term substitutes, cross-references, etc.) may be revised through the combination of a deletion entry and the reentry of the term set in its revised form.

Before processing any of the above transactions, the program will check for possible errors within the input – such as entering a duplicate term or deleting a non-existent term. Any inconsistencies in the input will be flagged in the Error Listing.

XMAIN3 – Print the Thesaurus File

Abstract

The Thesaurus File will be printed in the form of publishable copy or working copy. The printed Thesaurus will include authorized usage, substitutes, cross-references and scope notes for each term. The entire file, the subject section or the geographic section may be selected for listing. In addition, checkpoint restart is provided on an interrupt basis. Input for a new Permuted Thesaurus File may also be created.

Programmer's Notes

Publishable copy of the Thesaurus File. The nature of XMAIN3 is such that the publishable copy of the Thesaurus could not be produced from the basic program. To

produce publishable copy, there is a set of cards identified as the "Publications Patches to XMAIN3". When producing publishable copy of the Thesaurus File these cards should be inserted before the last card of the program deck and removed at the completion of the run.

The Permuted Thesaurus File. XMAIN3 was modified by URBANDOC to create punched card output of the term entries to be used as the input to XMAIN5. The nature of the program modification is such that when creating punched card output all terms are punched. It is not possible to punch some of the terms.

Unless the terms added to and deleted from the Thesaurus File have been accumulated on punched cards by the document analysts since the creation of the latest Permuted Thesaurus File, the entire file of terms must be recreated. For this option, the "Publication Patches to XMAIN3" must be used.

Report date. Report date is the date in the tape label for the file. For the Thesaurus File, this date is the date of the last file maintenance. For the Thesaurus Supplement File, it will be the date used in the selection of the terms (cut-off date).

XMAIN4 – Statistical Analysis of the Thesaurus File

Abstract

The structure of the Thesaurus will be examined for the number of terms, number of searchable terms, relationships of terms to each other in hierarchical and lateral relationships, etc. The Statistical Analysis will provide the frequency counts for both the subject and geographic sections of the Thesaurus File.

XMAIN5 – Permute the Thesaurus File

Abstract

When maintaining and printing the Permuted Thesaurus, an existing file may be updated or a new file created. Before the actual permutation of the words in the term, the entry is edited. Terms with words of excessive length will be listed for the document analyst's attention and will not be permuted for that word. A term containing a "stop word" also will not be permuted for that word. A final version of the Permuted Thesaurus may be printed in single or multiple copies. An output tape is created for subsequent updating.

Programmer's Notes

Concept of a word. The Thesaurus File sets a limit of fifty-four characters on the size of a term on the file. This does not imply a one-word term varying in length to fifty-four characters. For the most part, the longer terms in the URBANDOC Thesaurus contain several words which together are not more than fifty-four characters.

XMAIN5 will permute a term on the words which form it. For this program, a "word" begins with the first character after a blank and ends with the last character before the next blank. As such, a word cannot exceed twenty characters.

Stop word. A "stop word" is a user-supplied word which should not be used for permuting a term. These are also known as "common words" or "kill words", for example "an", "the", "to".

The permuted entry. The permutation process creates one entry for each word in a term, excluding "too long" terms and stop words. This principle is illustrated in the following sample of entries from the Permuted Thesaurus:

```
                BICYCLING
          BIDS, BIDDING
     BIDDING, BIDS
       SIGNS, BILLBOARDS
```

The entries are aligned along a center point. Each entry is positioned according to the word being permuted and its location within the term. The portion to the right of the alignment point is the "index entry" while the portion to the left is the "end of entry".

Although a term may not be longer than fifty-four characters on the Thesaurus File, extra positions have been allotted to these fields on the Permuted Thesaurus File to allow for the proper positioning of longer terms. The terms are fully formatted for the printed page on the tape file.

XMAIN6 – Create A Thesaurus Supplement File

Abstract

The Thesaurus Supplement was designed to reduce the number of new editions that must be published. The Thesaurus Supplement File will contain the new terms, subject, geographic, or both, that have been entered or revised after the cut-off date. A printed supplement can be obtained by using the Thesaurus Supplement File as the input to XMAIN3.

Programmer's Notes

Cut-off date. The cut-off date is a user supplied date which determines the selection of terms for the Thesaurus Supplement File. Any term will be included that has been entered or revised after the supplied date.

Date of the supplement. The dates of the Thesaurus File and the Thesaurus Supplement File are contained in the tape labels for the files. The date of the Thesaurus File is the date of the last file update. For the Thesaurus Supplement File, the date will be the cut-off date for the run.

20..1

Input Specifications

Thesaurus Input
(See Chapter VII, specifically Thesaurus Data Entry.)

a. *Main Term Format:*

Field	Size	Cols	Contents
Thesaurus Term	X(54)	1-54	
Entry Date	X(8)	55-62	mm/dd/yy
Filler	X(2)	63-64	
Change Date	X(8)	65-72	mm/dd/yy
Accession Number	X(4)	73-76	
Sequence Number	X	77	
Filler	X	78	
Truncation Code	X	79	
Entry Code	X	80	

b. *Substitute Term Format:*

Field	Size	Cols
Search Substitute	X(24)	1-24
Publications Substitute	X(35)	25-59
Filler	X(13)	60-72
Accession Number	X(4)	73-76
Sequence Number	X	77
Filler	X(3)	78-80

c. *Cross-Reference Format:*

Field	Size	Cols
Cross-reference Code	X	1
Filler	X(2)	2-3
Cross-reference	X(51)	4-54
Filler	X(18)	55-72
Accession Number	X(4)	73-76
Sequence Number	X	77
Filler	X(2)	78-79
Card Code	X	80

d. *Scope Note Format:*

Field	Size	Cols
Line	X(54)	1-54
Filler	X(18)	55-72
Accession Number	X(4)	73-76
Sequence Number	X	77

Filler	X(2)	78-79
Card Code	X	80

Permuted Thesaurus Input

a. *Term Card Format:*

Field	Size	Cols
Thesaurus Term	X(53)	1-53
Filler	X(27)	54-80

Tape File Specifications

Thesaurus Input File
Sorted Thesaurus Input File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	'1HDR'
Filler	X(4)	5- 8	
Label	X(16)	9-24	'DICTIONARY-INPUT'
Filler	X(6)	25-30	
Date	X(6)	31-36	mmddyy
Filler	X	37	
Reel Number	X(3)	38-40	

c. *Data Record Format:* Unblocked variable-length records, maximum size of 1296 characters

Field	Size	Cols	Contents
Record Size	X(3)	1- 3	
Term Size	X(3)	4- 6	
Cross-reference Size	X(3)	7- 9	
Substitution Flag	X	10	
Filler	X	11	
Truncation Flag	X	12	
Entry Code	X	13	
Term	X(54)	14- 67	
Entry Date	X(8)	68- 75	mm/dd/yy
Filler	X(2)	76- 77	
Change Date	X(8)	78- 85	mm/dd/yy
Search Substitute	X(24)	86- 109	
Publications Substitute	X(35)	110- 144	
Cross-references	X(1152)	145-1296	Occurs as a variable number of X(72) fields. Maximum of 16 occurrences.

d. *Trailer Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain 1EOR'.
Filler	X(36)	5-40	

Thesaurus File
Thesaurus Supplement File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	'1HDR'
Filler	X(5)	5- 9	
Label	X(15)	10-24	'DICTIONARY FILE'
Filler	X(6)	25-30	
Date	X(6)	31-36	mmddyy
Filler	X	37	
Reel Number	X(3)	38-40	

c. *Data Record Format:* Unblocked variable-length records, maximum size of 1296 characters

Field	Size	Cols	Contents
Record Size	X(3)	1- 3	
Term Size	X(3)	4- 6	
Cross-reference Size	X(3)	7- 9	
Substitution Flag	X	10	
Filler	X	11	
Truncation Flag	X	12	
Entry Code	X	13	
Term	X(54)	14- 67	
Entry Date	X(8)	68- 75	mm/dd/yy
Filler	X(2)	76- 77	
Change Date	X(8)	78- 85	mm/dd/yy
Search Substitute	X(24)	86- 109	
Publications Substitute	X(35)	110- 144	
Cross-references	X(1152)	145-1296	Occurs as a variable number of X(72) fields. Maximum of 16 occurrences.

d. *Trailer Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(36)	5-40	

Permuted Thesaurus File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

Data Record Format: Fixed-length records of 80 characters
Blocking factor of 6
Padding record of 9s

Field	Size	Cols	Contents
End of Entry	X(34)	1-34	)See XMAIN5,
Index Entry	X(35)	35-69	)Programmer's Notes
Filler	X(11)	70-80	

References to the URBANDOC Final Report

Much of the information presented in this chapter is designed to be used with sections of the *General Manual* (G.M.) and other sections of the *Operations Manual* (O.M.)

For additional information on the Thesaurus File, its format and the considerations in creating and maintaining entries on the file, see:

Manual	Chapter	Section
G.M.	IV – Document Analysis: Content	Thesaurus
G.M.	VI – Systems Modules: Input	Thesaurus Module: Thesaurus File Description
O.M.	VII – Data Entry	Thesaurus Data Entry
O.M.	X – Error Listings and Systems Messages	Thesaurus Module

For additional information on the design and goals of the Thesaurus Module, see:

Manual	Chapter	Section
G.M.	VI – Systems Modules: Input	Thesaurus Module: Function, Tasks

For additional information on operating this portion of the system see:

Manual	Chapter	Section
O.M.	VIII – Processing Cycles	Input Processing Cycle, Miscellaneous Thesaurus Products
O.M.	IX – Operating Instructions	Thesaurus Module
O.M.	XI – Tape Library and Report Controls	Thesaurus Module
O.M.	XII – Timing	Thesaurus Module

PRE-EDIT MODULE

Program Inventory

Program Number	Program Function	Source	Basic or On-demand*	Programming Language	Configuration Set**
E0010	To format records to the requirements of the Pre-edit and File Maintenance Modules	URBANDOC	Basic	COBOL	Either 1 or 2
E0020	To expand the Document Input File, generate all the standard input units and list the input	URBANDOC	Basic	COBOL	Either 1 or 2
E0030	To edit the Document Input File for such errors as sequence, coding, format, etc., and list the detected errors	URBANDOC	Basic	COBOL	1
E0040	To update the original Document Input File with the contents of the Document Revision File, making additions, deletions and changes	URBANDOC	Basic	COBOL	Either 1 or 2

*Basic: Part of the standard processing of the Document Input File.

**Set 1: 12K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; 1311 disk drive; advanced programming; high-low-equal compare; sense switches.

Set 2: 8K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; advanced programming; high-low-equal compare; sense swithces.

Program Abstracts

E0010 – Card-To-Tape of the Document Input

Abstract

E0010 will create a tape image Document Input File in the format required by the Pre-edit and File Maintenance Modules. Advantage will be taken of the maximum blocking factor allowed by CFS. Some sequence checking will be performed.

Programmer's Note

Revisions to the input. This program can also process the corresponding Document Revision File. (See Chapter VII, specifically Document Data Entry. See also Chapter VIII, specifically the Editing and Validation Cycle and Input Processing Cycle).

E0015 – Sort the Document Input File

Abstract

The Document Input File is created in the same sequence as the Document Input. Since the Document Master File will be updated with this input, the Document Input File must be sorted to document number and unit number sequence.

Programmer's Notes

Revisions to the input. This program can also process the corresponding Document Revision File. (See Chapter VII, specifically Document Data Entry. See also Chapter VIII, specifically the Editing and Validation Cycle and Input Processing Cycle.)

Sequence of the Document Input File. The total input file will be sorted to document number sequence by the use of the document number field in each record. The unit number field contains a sequential number which serves to keep the individual records within the reference in the correct sequence. (See Chapter VII, specifically Document Data Entry.)

E0020 – Expand and List the Document Input File

Abstract

The Document Input File will be expanded to the full format required by the CFS system. The program will generate such information as the heading units for a reference, the additional geographic entries for SMSA and city size, the subject headings and the imprint and collation information for subdocuments. The program will also expand to CFS format the date entries and subdescriptor entries. A formatted listing of the input or output may also be produced.

Programmer's Notes

Expansion of the input. In an attempt to maximize the rate of creating document input

while minimizing the data entry costs, URBANDOC devised a set of basic input to be entered. (See Chapter VII, specifically Document Data Entry. See also the *General Manual,* Chapter VI, specifically the Pre-edit Module.) E0020 was designed to expand this basic data to the format of the CFS system by generating additional entries and completing certain fields in other entries.

Revision to the input. Because E0020 always generates the additional entries for the reference, this program cannot be used during the Input Processing Cycle. (See Chapter VIII, specifically the Input Processing Cycle.) If the program was used, some problems that could arise are the duplication of entries and the accidental deletion of desired information.

The original purpose of E0020 was to reduce the amount of effort on the part of the document analyst in entering the data into the system. Since this program cannot be used for revisions, this data must be entered according to CFS specifications. (See Chapter VII, specifically Editing and Validation.)

Reference to an index journal issue. It is possible to insert, as a descriptor, the issue number of the *Input Index* in which the reference appears. (See Chapter IX, specifically the operating instructions for E0020.) Since this particular technique has been used only in test situations, its actual utility is yet to be determined. It could be of help in creating cumulative issues for the *Input Index.* (See Chapter VIII, specifically Miscellaneous Publications Products.)

E0030 – Edit the Document Input File

Abstract

The Full Document Input File will be edited for errors in sequence, coding, format, authorized bibliographic elements, etc. The detected errors will be printed with an error code as part of the Pre-edit Listing.

Programmer's Notes

Terminal punctuation for geographic terms. Originally, a terminal punctuation character of either '.', '*', or '/' was required for each geographic term for which a numeric substitution is generated. This check was accordingly built into the system. When terms were added that have no numeric substitution, no punctuation character was specified as the last character of the term, e.g. "Miami Valley Ohio Region". However, the program continued to check for the terminal punctuation character for these terms. When terminal punctuation is missing, an error message will occur even though the term may be correct. For these cases, ignore the message.

More than nine errors. The system based its edit checking on the assumption that no more than nine errors would occur within any one unit. If a unit should contain more than nine errors, the additional ones will not be detected.

Valid bibliographic elements. E0030 is one of the few programs that checks element number by a pre-programmed list rather than through a lead card, that is a program control card providing data required for processing. If the authorized bibliographic element numbers are changed, the program must be modified and recompiled or incorrect messages will result.

E0040 – Revise the Document Input File

Abstract

The Full Document Input File will be revised by the Document Revision File. Entire references or units of a reference (one or more of the logical records) may be added or deleted. Individual units may be replaced. A listing of the processed transactions will be printed with codes for the action taken. Non-processed units will also be listed.

Input Specifications

Document Input
(See Chapter VII, specifically the Document Worksheet)

a. *Descriptor Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Expand Code	X(2)	21-22
Pre-code	X	23
Descriptor	X(35)	24-58
First Subdescriptor	X(12)	59-70
Filler	X(10)	71-80

b. *Subdescriptor Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(14)	21-34
Additional Subdescriptor	X(12)	35-46
Filler	X(34)	47-80

c. *Dates Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	

Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Filler	X(2)	21-22	
Pre-code	X	23	
Descriptor	X(5)	24-28	'#D ATES'
Filler	X	29	
Publications Date	X(4)	30-33	yyyy
Filler	X	34	
Entry Date	X(6)	35-40	yyyymm
Filler	X	41	
Content Date	X(4)	42-45	yyyy
Filler	X	46	
Content Date	X(4)	47-50	yyyy, used only for ranges
Filler	X(30)	51-80	

d. *Bibliographic Data Format:*

Field	**Size**	**Cols**
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Bibliographic Line	X(58)	23-80

Document Revisions
(See Chapter VII, specifically Document Master File Revisions)

a. *Header Format:*

Field	**Size**	**Cols**
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(2)	21-22
Data Entry Date	X(24)	23-46
Filler	X(34)	47-80

b. *Descriptor Format:*

Field	**Size**	**Cols**	**Contents**
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Filler	X(2)	21-22	
Pre-code	X	23	

Descriptor	X(35)	24-58	
First Subdescriptor	X(12)	59-70	
Numerical Value	X(10)	71-80	Used only for publications date

c. *Subdescriptor Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Descriptor Tie	X(2)	21-22	
Subdescriptor Number	X(2)	23-24	
Descriptor Root	X(6)	25-30	
Filler	X(4)	31-34	
Subdescriptor	X(12)	35-46	Used only for content dates or entry date
Numerical Value	X(10)	47-56	Used only for content dates or entry date
Filler	X(2)	57-58	
Subdescriptor	X(12)	59-70	Used only for content dates
Numerical Value	X(10)	71-80	Used only for content dates

d. *Bibliographic Data Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Bibliographic Line	X(58)	23-80

e. *Subject Heading Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Type	X	19
Entry Code	X	20
Element Number	X(2)	21-22
Subject Code	X	23
Subject Sequence	X	24
Subject Heading	X(35)	25-59
Filler	X(21)	60-80

f. *Delete Reference Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Filler	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(60)	21-80

Tape File Specifications

Document Input File
Sorted Document Input File

a. *File Format:*

No header label
Data records, tape mark
No trailer label

b. *Data Records Format:*

Fixed-length records of 84 characters
Blocking factor of 10
Padding record of 9s

Five data records:
- Descriptor
- Subdescriptor
- Dates
- Bibliographic Data
- Delete Reference

c. *Descriptor Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Expand Code	X(2)	21-22
Pre-code	X	23
Descriptor	X(35)	24-58
First Subdescriptor	X(12)	59-70
Filler	X(13)	71-83
Record Mark	X	84

d. *Subdescriptor Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(14)	21-34
Additional Subdescriptor	X(12)	35-46
Filler	X(37)	46-83
Record Mark	X	84

e. *Dates Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Filler	X(2)	21-22	
Pre-code	X	23	
Descriptor	X(6)	24-29	
Publications Date	X(4)	30-33	yyyy
Filler	X	34	
Entry Date	X(6)	35-40	yyyymm
Filler	X	41	
Content Date	X(4)	42-45	yyyy
Filler	X	46	
Content Date	X(4)	47-50	yyyy, used only for ranges
Filler	X(33)	51-83	
Record Mark	X	84	

f. *Bibliographic Data Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Bibliographic Line	X(58)	23-80
Filler	X(3)	81-83
Record Mark	X	84

g. *Delete Reference Format:*

Field	Size	Cols
Document Number	X(14)	1-14

Filler	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(63)	21-83
Record Mark	X	84

Full Document Input File
Document Revision File
Sorted Document Revision File

a. *File Format:*

No header labels
Data records, tape mark
No trailer labels

b. *Data Records Format:*

Fixed-length records of 84 characters
Blocking factor of 10
Padding record of 9s

Five data records:
- Header
- Descriptor
- Subdescriptor
- Bibliographic Data
- Subject Heading

c. *Header Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Filler	X(2)	21-22
Data Entry Date	X(24)	23-46
Filler	X(37)	47-83
Record Mark	X	84

d. *Descriptor Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Filler	X(2)	21-22	
Pre-code	X	23	
Descriptor	X(35)	24-58	

First Subdescriptor	X(12)	59-70	
Numerical Value	X(10)	71-80	Used only for publications date
Filler	X(3)	81-83	
Record Mark	X	84	

e. *Subdescriptor Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Descriptor Tie	X(2)	21-22	
Subdescriptor Number	X(2)	23-24	
Descriptor Root	X(6)	25-30	
Filler	X(4)	31-34	
Subdescriptor	X(12)	35-46	Used only for content dates or entry date
Numerical Value	X(10)	47-56	Used only for content dates or entry date
Filler	X(2)	57-58	
Subdescriptor	X(12)	59-70	Used only for content dates
Numerical Value	X(10)	71-80	Used only for content dates
Filler	X(3)	81-83	
Record Mark	X	84	

f. *Bibliographic Data Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Bibliographic Line	X(58)	23-80
Filler	X(3)	81-83
Record Mark	X	84

g. *Subject Heading Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Type	X	19
Entry Code	X	20
Element Number	X(2)	21-22
Subject Code	X	23
Subject Sequence	X	24

Subject Heading	X(35)	25-59
Filler	X(24)	60-83
Record Mark	X	84

References to the URBANDOC Final Report

Much of the information presented in this chapter is designed to be used with sections of the *General Manual* (G.M.) and other sections of the *Operations Manual* (O.M.)

For additional information on the Document Input File, its format and the considerations in creating and maintaining the entries on the file, see:

Manual	Chapter	Section
G.M.	I – Introduction	The Bibliographic Records
G.M.	II – Document Identification	URBANDOC Document Numbers: General
G.M.	III – Document Analysis: Descriptive	General Considerations
G.M.	IV – Document Analysis: Content	General Considerations

For more information on the design and goals of the Pre-edit Module, see:

Manual	Chapter	Section
G.M.	VI – Systems Modules: Input	Pre-edit Module: Function, Tasks

For more information on operating this portion of the system, see:

Manual	Chapter	Section
O.M.	VIII – Processing Cycles	Editing and Validation Cycle, Input Processing Cycle
O.M.	IX – Operating Instructions	Pre-edit Module
O.M.	XI – Tape Library and Report Controls	Pre-edit Module
O.M.	XII – Timing	Pre-edit Module

FILE MAINTENANCE MODULE

The Document Master File

The Document Master File is composed of references to individual documents. Each reference is entered as a series of individual tape records, each record containing a specific, integral unit of information. The individual tape record will be referred to as a "record". "Reference" will apply to the aggregate of all tape records for a document.

The primary uses of the Document Master File are in the identification, isolation and presentation of references. Some of the entered information describes and defines the contents; this portion, subject to interrogation in a search, is called the Searchable Information. The remaining data is useful in the presentation of the document reference and during a manual bibliographic search. It is not used during a computer search. This information is called Free Text Information. In addition to these two types, the system also uses a third, Program Utility Information, to facilitate computer processing during file maintenance and search. Program Utility Information is internal to the programming system and is not made available to the user.

The Searchable Information or the Searchable Record is the first tape record of the reference. The Free Text Information is stored as a variable number of Free Text Records with a maximum of ninety-nine records. The Searchable Record is composed of the Document Master Identification, Program Utility Data and Searchable Data. The Free Text Record is composed of the Document Master Identification and Free Text Data. (See Figure 3.)

The Document Master Identification contains the document identification number, segment number, and date of data entry. The document number and date of data entry is the same for all records of a reference. The segment number is a three position code identifying the contents of the record. For a Searchable Record, the code is '000'. For Free Text Records, the code is that of the bibliographic element entered in the record. (These codes are listed later in the section.)

The Program Utility Data in the Searchable Record is for purposes of program efficiency during search. This section immediately follows the Document Master Identification in the record. The Program Utility Data is a table of pointers, with one entry for each descriptor assigned to the reference. Each pointer is eleven characters giving information about the location and nature of the descriptors (See Figure 3.):

Address: a three-position field containing the displacement (from the first position of the individual tape record) of the descriptor within the Searchable Record. For those descriptors with a displacement greater than 999, the address is represented in the 1401 machine-language code for core storage addresses (See Figure 2.);

Rank: a three-position field indicating the descriptor's sequence in the reference. For example, if the descriptor were the first assigned, rank would be '001';

Size: a one-position code for the length of a descriptor. Each descriptor is stored as

either a fixed-length field of twelve characters (size 1) or twenty-four characters (size 2) (See Chapter VII, specifically Document Data Entry.);

The number of modifiers assigned to each descriptor. If a descriptor has no modifiers, this field is zero;

Type code: the descriptor's status, '*' for precise, '#' for common and '-' for internal.

While the descriptors are stored in the order in which they were assigned, the pointers are arranged alphabetically by descriptor to increase the efficiency of the search process. The end of the Program Utility Section is designated by a dummy pointer of three characters coded 'END'.

The descriptors and modifiers immediately follow the 'END' constant of the pointer table. The number of descriptors assigned to a record cannot exceed ninety-nine. The number of modifiers assigned to one descriptor cannot exceed ninety-nine. However, the overriding limit on the Searchable Record is that the combined length of all the sections of the record cannot exceed one tape record of 2200 characters. (See Figure 3.)

The descriptor is the basic element of the search. All descriptors within a single record are considered to be independent and of parallel significance. The modifier, when assigned, represents an elaboration or refinement of a descriptor. It is in a dependent relationship to the descriptor to which it is assigned. Modifiers relating to the same descriptor are considered to be logically independent of each other.

A descriptor may be variable in length to a maximum of fifty-four characters when stored on the Thesaurus File. (See Chapter VII, specifically Thesaurus Data Entry. See also *General Manual*, Chapter VI, specifically the Thesaurus File Description. When stored on the Document Master File, each descriptor is a fixed-length field of either eleven or twenty-three characters. Smaller entries will be padded with blanks to conform to the eleven or twenty-three character length specification. The fixed-length field was adopted for programming and tape storage efficiency. (See Chapter VII, specifically Document Data Entry.)

The modifier is used to make more specific use of a descriptor. This may be done either through a natural language subdescriptor or a value (numerical subdescriptor). Either or both portions of the modifier may be used. Regardless of the portion completed, the modifier is stored as a fixed-length field of twenty-one characters.

The natural language subdescriptor is a secondary-level descriptor. (See Chapter VII, specifically Thesaurus Data Entry. See also *General Manual,* Chapter VI, specifically the Thesaurus File Description. In the Document Master File it is stored as a fixed-length field of twelve characters within the modifier. A pre-code is not used with this field. Like the descriptor, longer terms will be truncated. Shorter terms will be padded with blanks to twelve characters. (See Chapter VII, specifically Document Data Entry.)

The numerical value, in the CFS system, is a variation of the scientific notation for

decimal numbers. Each number or value is represented through eight characters of numerical data. Each value is composed of a six-position fraction and a two-position exponent representing a power of 10. For example: the number 1970 would consist of the fraction .197000 and an exponent of 04. That is, the number 1970 could be obtained by multiplying the fraction .197000 by 10,000 (10 to the fourth power). CFS also makes provision for a sign character, indicating whether the number has a positive or a negative value. For URBANDOC's purposes, numbers are always positive.

Numerical values may range from an upper to a lower limit. To indicate whether a value is a single value or the upper or lower limit of a range, a letter code is included as part of each modifier. All modifiers must immediately follow the parent descriptor in the record. (See Chapter VII, specifically Document Data Entry.)

All modifiers relate directly to the descriptor. If the subdescriptor portion of a modifier is present, the numerical value applies directly to the subdescriptor. If the subdescriptor is not used, the value applies to the descriptor itself. URBANDOC uses a numerical value only with date analysis descriptors. However, if used eleswhere, the same rules apply.

The Free Text Section contains the bibliographic information about a document reference. The organization of this information is by element, for example, author, title, imprint and collation, subject headings, etc. Within each element of bibliographic information, the data is neither reformatted nor edited (except for valid element number).

The Free Text Information is divided into tape records (or segments) with one record corresponding to one bibliographic element. The CFS system provided for ninety-nine elements; except for one record, it placed no restrictions on the use of these records by the document analyst and/or systems analyst. The one identified segment was '99' to be used for subject headings. The purpose of segmentation is to permit the selection and printing of relevant portions of bibliographic information, thus avoiding the unnecessary recall of the entire reference. (See Figure 3.)

URBANDOC did not identify ninety-nine bibliographic elements which it wished to include. It identified a total of twenty-five bibliographic elements which could be entered:

01	personal author
02	corporate author
03	anonymous
04	acronyms
05	joint personal author
06	joint corporate author
07	consultant
08	miscellaneous corporate name
09	miscellanecus local place name
10	corporate author name cross-reference
11	distinctive title
13	distinctive series title

15 non-distinctive title
16 French title, if available
17 German title, if available
18 Spanish title, if available
21 imprint
22 imprint for a subdocument
23 abstract and notation of content
25 UPAP project number
31 statutory citation
36 literature citation
56 acquisition information
97 geographic index name
99 subject heading

The element numbers used by URBANDOC are not consecutive. Gaps exist to add new elements in a particular sequence or because previously used elements were discontinued. The sequence of bibliographic elements is significant since this information is always retrieved from the Document Master File in element number sequence.

The size of the total Free Text Information is a function of the number of elements assigned and the maximum size for each element's entry. There may be a maximum of ninety-nine elements of bibliographic information. Each element (including the Document Master Identification) may be no more than one tape record (2200 characters). Therefore, the total Free Text Information may be no more than 99 x 2200 characters (217,800 characters).

The bibliographic information begins immediately after the Document Master Identification. This corresponds to the first character in the pointer table of the Searchable Record. Each line of bibliographic information entered is given a terminal punctuation character of a record mark (0-2-8) by the CFS system. This character must be included when calculating the size of a bibliographic element.

The Inverted File

The Inverted File is a series of records for each descriptor used by the system. Each individual record dealing with the usage of one descriptor is composed of two sections: the Identification Section and the Document Number List.

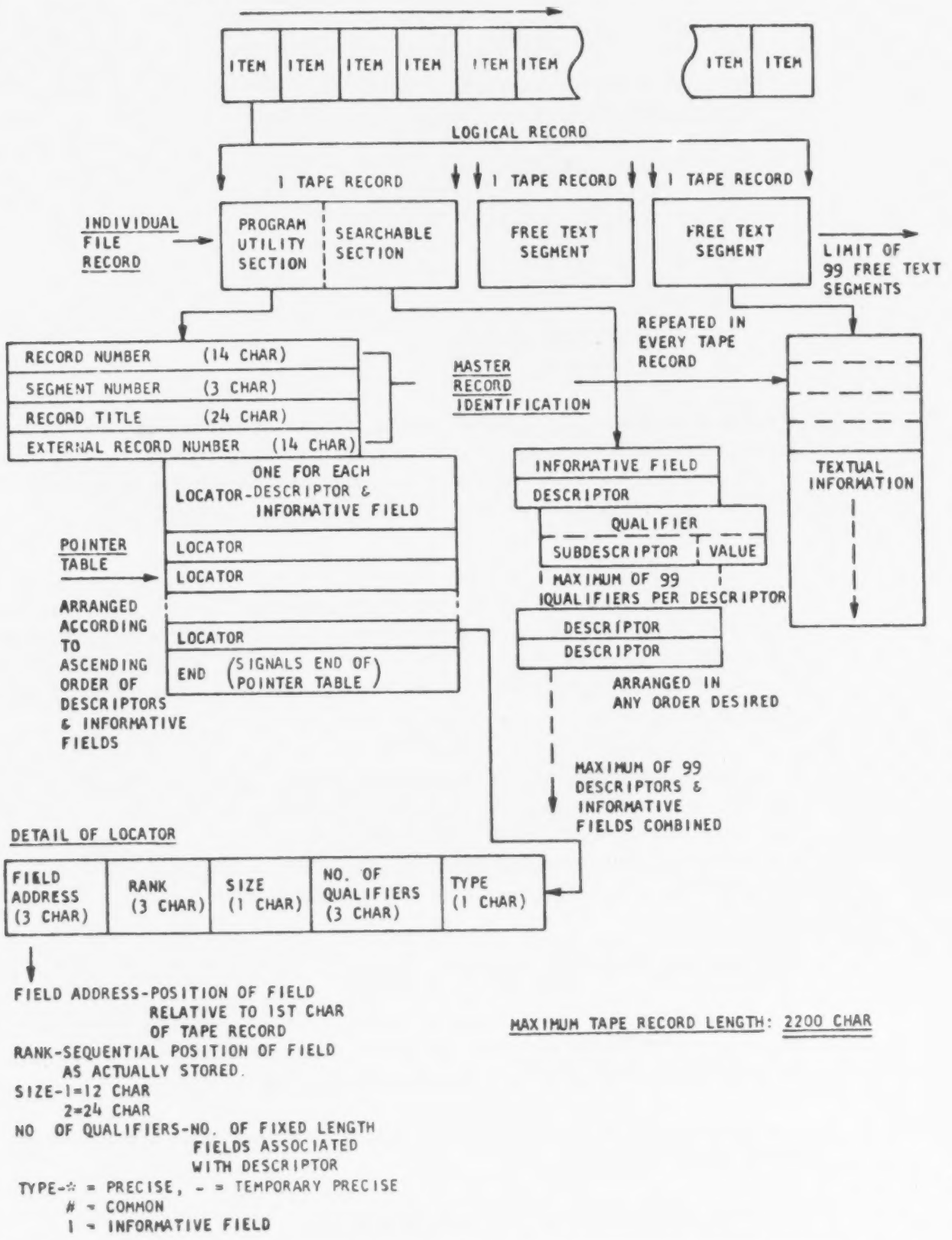
The Identification Section consists of a series of fields which serve to define the way the term has been used:

Type: precise, common or internal descriptor;

Frequency count: number of documents in which the term has been used as a descriptor;

Date of last addition to the file as a descriptor;

DOCUMENT MASTER FILE RECORD LAYOUT[1]



[1] International Business Machines, Data Processing Division, *1401 Information Storage and Retrieval System. (The Combined File Search System)* by Donald Prentice, Gary de Graw, Alice Smith and I. Albert Warheit, Rev. III, 1 v. (White Plains, N.Y., 1966), p.1.08.

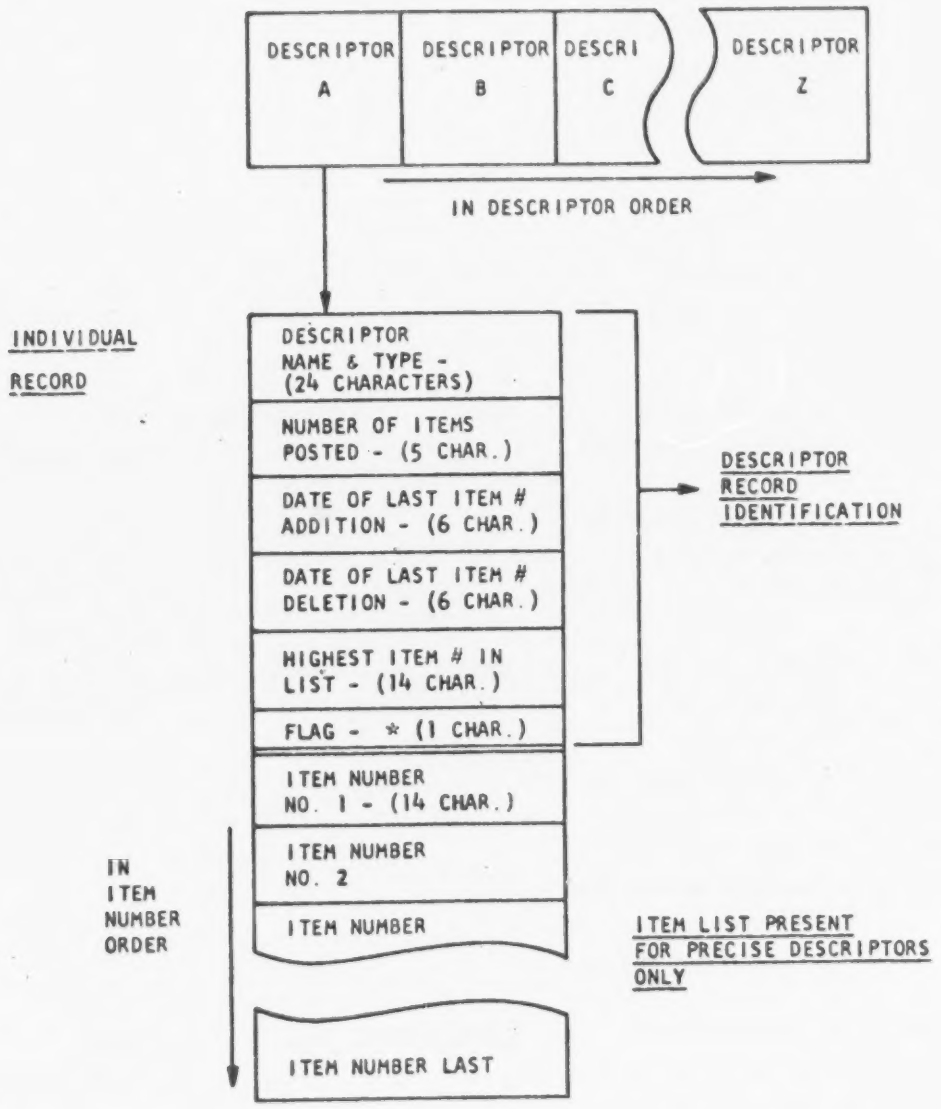
Figure 3

Date of last deletion from the file as a descriptor; and

Highest document number in the file (used only by the system when searching).

The Document Number List consists of a variable number of fourteen-character fields containing the document numbers in which the term has been used as a descriptor. The maximum number of entries within one record is one hundred and fifty. If more than this number of postings occur, there will be a continuing record on the Inverted File for this descriptor.

INVERTED FILE RECORD LAYOUT[2]


DESCRIPTOR A
DESCRIPTOR B
DESCRI C
DESCRIPTOR Z
IN DESCRIPTOR ORDER
INDIVIDUAL RECORD
DESCRIPTOR NAME & TYPE - (24 CHARACTERS)
NUMBER OF ITEMS POSTED - (5 CHAR.)
DATE OF LAST ITEM # ADDITION - (6 CHAR.)
DATE OF LAST ITEM # DELETION - (6 CHAR.)
HIGHEST ITEM # IN LIST - (14 CHAR.)
FLAG - ☆ (1 CHAR.)
DESCRIPTOR RECORD IDENTIFICATION
ITEM NUMBER NO. 1 - (14 CHAR.)
ITEM NUMBER NO. 2
ITEM NUMBER
IN ITEM NUMBER ORDER
ITEM LIST PRESENT FOR PRECISE DESCRIPTORS ONLY
ITEM NUMBER LAST

[2] Ibid., p 1.12.

Figure 4

Program Inventory

Program Number	Program Function	Source	Basic or On-demand*	Programming Language	Configuration Set**
MAINT1	To format the input for CFS specifications and perform the final sequence check	CFS	Basic	Tape AUTOCODER	Either 1 or 2
MAINT2	To validate the descriptors, sub-descriptors and subject headings headings against the Thesaurus File for such factors as authorized terms, usage of terms, etc.	CFS	Basic	Tape AUTOCODER	Either 1 or 2
MAINT3	To print the CFS Edit Listing	CFS	Basic	Tape AUTOCODER	Either 1 or 2
MAINT4	To update the Document Master File	CFS	Basic	Tape AUTOCODER	Either 1 or 2
MAINT5	To provide tape labelling instructions for output of MAINT4 and set-up instructions for MAINT6	CFS	On-demand	Tape AUTOCODER	Either 1 or 2
MAINT6	To print Master File Activities and create the Descriptor Input File	CFS	Basic	Tape AUTOCODER	Either 1 or 2
MAINT7	To copy the Document Master File or recreate the Descriptor Input File				
MAINT8	To generate and print statistics for the contents of the Document Master File	URBANDOC	Basic	Tape AUTOCODER	Either 1 or 2

Program Number	Program Function	Source	Basic or On-demand*	Programming Language	Configuration Set**
MAINT9	To generate the Document Master Subset File	URBANDOC	On-demand	Disk AUTOCODER	Either 1 or 2
LIBPRT	To print a listing of the Document Master File	CFS	On-demand	Tape AUTOCODER	Either 1 or 2
DMAIN1	To create the Inverted File and check for certain types of errors	CFS	Basic	Tape AUTOCODER	Either 1 or 2
DMAIN2	To produce the Summary Listing of the Inverted File for an overview of each term including frequency, type, etc.	CFS	Basic	Tape AUTOCODER	Either 1 or 2
DMAIN3	To produce the Detail Listing of the usage of terms on the Inverted File with postings of the document numbers in which the terms have been used	CFS	Basic	Tape AUTOCODER	Either 1 or 2

* Basic: Part of the standard processing when updating the Document Master File.
On-demand: Performed only upon request.

**Set 1: 12K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; 1311 disk drive; advanced programming; high-low-equal compare; sense switches.

Set 2: 8K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; advacnced programming; high-low-equal compare; sense swithces.

Program Abstracts

MAINT1 — Format the Document Input File

Abstract

The first step in the file maintenance procedures is to format the Document Input File according to CFS specifications. This includes the final verification of the sequence of the input by document number and unit number. Out-of-sequence units are listed and removed from further processing. Among the valid input, document numbers, descriptors, subdescriptors and subject headings are reduced to their most compact form. Two output files are created — the formated valid input (Docu-to-Tape File) and all terms to be validated against the Thesaurus File (Look-Up Input File).

Programmer's Notes

Compaction of descriptors, subdescriptors and subject headings. MAINT1 scans the input terms for blanks within the field. Any leading blanks or embedded blanks (except for the word separator) in the term will be removed. The intellectual combination of descriptors and subdescriptors will not be affected by the compaction process since it operates independently on each entry. Multiple words within terms are allowed as long as the length limitation is not exceeded. Terms will be compacted to the extent that only one blank space is left between words. The compacted term will be stored left-justified in the field to agree with the terms as stored on the Thesaurus File. (See Chapter II, specifically the Thesaurus File and XMAIN1.)

Compaction of document numbers. MAINT1 will scan each document number and remove any leading or embedded blanks in the field. After the removal of the leading or embedded blanks, the document number will be shifted left in the field. The removal of these blanks can create serious problems for both processing and file organization. First, the embedded blanks have been included to maintain consistency in document number form within a series. (See *General Manual,* Chapter II, specifically URBANDOC Document Numbers: By Series.) For example, consider the document number 'DRAA LP66 391'. As a result of MAINT1, it would appear 'DRAALP66391' on the Edit File and 'DRAA LP66 391' on the Docu-to-Tape File and the two tapes could not be match-purged of errors. (See abstracts of SORT1, MAINT2, SORTM2, MAINT3.)

URBANDOC has avoided this problem by using a filler character of '.' in place of the blank to prevent the document numbers from being changed. No filler character is necessary to complete the end of the field, only to reserve places within the field.

In the event that this error condition did occur, the processing cycle would have to be cancelled, the document number in error corrected and processing restarted from the beginning.

Deleting descriptors, subdescriptors and subject headings. When revising the Document Master File, it may be necessary to delete descriptors, subdescriptors and subject headings from references already on the Document Master File. MAINT1 would include these

terms on the Look-Up Input File for validation against the Thesaurus File. If these terms have been removed from the Thesaurus File the deletion entries must be processed as a separate revision cycle in which term validation is bypassed. (See abstract and programmer's notes on MAINT3.)

SORTM1 -- Sort the Look-Up Input File

Abstract

Since the Look-Up Input File (descriptors, subdescriptors and subject headings) will be validated against the Thesaurus File, it must be sorted to the same sequence. The Look-Up Input File is created in document number sequence. At the conclusion of the sort, the file will be in term sequence.

MAINT2 -- Validate the Look-Up Input File

Abstract

The terms on the Look-Up Input File are validated against the Thesaurus File for four conditions. (1) Is the term an authorized one? (2) Has the term been used correctly as a descriptor or subdescriptor? (3) Is the coding of the term correct? (4) Is there a preferred form of the term other than the one used in the input?

A term failing any of the first three tests is considered an invalid term. A term qualifying for the fourth condition is considered to have a substitute. It is here that the change from place name to geographic code is made. All other terms are valid for use in their input form. Only the invalid terms and the substitutes are included as part of the Bad Look-Up File. Along with each term on the Bad Look-Up File is a code indicating the reason for its inclusion, either a preferred term or the specific error.

Programmer's Note

The publications substitute. The CFS system provides two substitutes for each thesaurus term: a search substitute for content analysis and a publications substitute for subject headings. URBANDOC conventions require that the analyst enter the preferred form during document analysis. As a result, only one substitute was entered into the Thesaurus File. Care must be taken to use the preferred form for both descriptors and subject headings during document analysis. If the original term is used for a subject heading, it will be replaced by blanks during the validation process since there are no subject heading substitutes on the Thesaurus File.

SORTM2 -- Sort the Bad Look-Up File

Abstract

The Docu-to-Tape File will be revised according to the contents of the Bad Look-Up File. Since the Bad Look-Up File is created in term sequence, it must be sorted to document number and unit number sequence to agree with the Docu-to-Tape File. This file will update the contents of the Docu-to-Tape File as part of MAINT2 processing.

MAINT3 – Edit the Document Input File

Abstract

MAINT3 is the major edit step in the CFS system; it updates the input with the results of the Thesaurus File validation and edits the data according to CFS specifications for code verification, sequence of bibliographic elements, and combinations of input units.

All invalid units are removed from further processing and printed as part of the Edit Listing. Substitutes remain in the input but are listed in the Edit Listing for audit trail purposes. Part of the Edit Listing is an error code to facilitate correction of the units by the document analysts. (See Chapter X, specifically MAINT3.) If the unit is acceptable because it has not failed any edit check, the unit is reformatted (for sake of program efficiency) for the actual master file update.

Programmer's Notes

Disposition of Error Codes 30-32. This comment applies only to the processing of revisions to the Document Master File. If any errors are detected within the value portion of the modifiers, the descriptor and its modifiers will not be processed. Although the error message seems to indicate that only the modifiers were non-processed, both the descriptor and the modifiers must be reentered. For URBANDOC, this problem only occurs for descriptors used in data analysis; the value portion is not used with other descriptors.

Spacing of the Edit Listing. When processing new input, the messages are grouped together by document number. This is not the case for revisions. In this instance, the spacing between documents is erratic; sometimes error messages for several documents will be grouped together.

Bypassing term validation. MAINT3 offers the user the ability to enter all information directly onto the Document Master File and Inverted File without validating descriptors, subdescriptors and subject heading, or without making computer substitutions of one form for another (as with the geographic descriptors). The user may start file maintenance processing directly with MAINT3, bypassing MAINT1, SORTM1, MAINT2 and SORTM2. (See Chapter IX, specifically the operating instructions for MAINT3.) URBANDOC did not use the direct-to-master-file option as part of its processing. However, this procedure might be necessary to remove descriptors, subdescriptors and subject headings from the Document Master File and Inverted File once the authorized entry for the term has been removed from the Thesaurus File.

MAINT4 – Update the Document Master File

Abstract

The Document Master File will be updated with the contents of the Edit File (the valid and formatted input). Transactions may include the addition or deletion of entire

document references, addition or deletion of free text segments (bibliographic elements) or changes to the searchable segment (content analysis) or free text segment '99' (subject headings). Certain systems conditions such as exceeding the record size or duplication of the document number would prevent a transaction from being processed. If any such condition is detected, the non-processed record is given a reason code to be included in an error listing. (See Chapter X, specifically the section on MAINT6.) Any descriptors are flagged for the update of the Inverted File.

Programmer's Notes

Inserting descriptors by rank. The CFS documentation states that when revising a reference, descriptors may be added either at the end of the Searchable Record or inserted into place by rank within the record. URBANDOC discovered a programming error when inserting descriptors by rank. The transaction is processed without any error messages. However, an incorrect pointer and meaningless descriptor is stored on the Document Master File and Inverted File. New descriptors should be added to the end of the searchable record; no rank should be specified. (See Chapter VII, specifically the section on Document Master File Revisions.) It was not reasonably possible to correct this complex CFS error.

Adding references with duplicate document numbers. The CFS documentation states that duplicate document numbers will not be processed. Actually, when a duplicate document number is encountered, the descriptive analysis data and the non-repetitive descriptors and subdescriptors will be added to the existing reference. That reference must be removed from the Document Master File and both references reentered with unique identification numbers.

MAINT5 – Print On-Line Operator Messages

Abstract

MAINT5 is an intermediary program which performs no actual processing of the input data but which serves as a control program. It provides the operator with instructions for labelling the tape output of MAINT4 and for setting up MAINT6. The same information is provided in the operating instructions. (See Chapter IX, specifically the operating instructions for MAINT4, MAINT5, MAINT6.)

MAINT6 – Print Update of the Document Master File

Abstract

MAINT6 will print the update of the Document Master File and create the input to the Inverted File. The transactions to the Document Master File may be printed in their entirety, by class, or not at all. If printed by class, the subsets of all additions, all deletions, and/or non-processed changes may be selected. Each subset is mutually exclusive, but multiple subsets may be printed. For the non-processed changes, the error

code is also printed. (See Chapter XI, specifically the File Maintenance Module.) The transactions involving descriptors are separated from the rest of the bibliographic input as the Descriptor Input File. This file is used to update the Inverted File.

Programmer's Notes

Listing non-processed changes. On certain occasions, a processing error will occur when the operator tries to list the non-processed changes to the Document Master File (Sense Switches A and D). The first attempt to remedy the error should be to restart the report through Check-Reset, Start-Reset and Start. If this does not correct the problem, restart MAINT6 bypassing listing the transactions and proceed directly to the non-processed changes. (See Chapter IX, specifically the operating instructions for MAINT6.)

Differentiation between descriptors and subdescriptors. The descriptors being added to or deleted from the references on the Document Master File are further processed for inclusion on the Inverted File where the user can obtain such information about the descriptor as frequency and place of usage. These results occur only for descriptors. No further processing occurs for subdescriptors. Once added to the Document Master File, subdescriptors can only be accessed through a computer search in conjunction with the descriptor which they modify.

MAINT7 — Process the Document Master File

Abstract

MAINT7 is not one of the programs that is used during normal file maintenance but provides additional support for the Document Master File and Inverted File. All processing handled by MAINT7 is based on either the reproduction of the Document Master File or the recreation of the Descriptor Input File for a new Inverted File. As independent actions, either the Document Master File may be copied or a new Descriptor Input File created. By combining these capabilities, two Document Master Files may be merged and a combined Descriptor Input File created. Checking is done to be sure that duplicate document numbers do not exist and the files to be merged are in sequence.

MAINT8 — Statistical Analysis of the Document Master File

Abstract

While the file maintenance procedures provide a count of the number of documents in the Document Master File, other statistics are needed. The listings of the Inverted File furnish a more detailed analysis of the use of the descriptors. MAINT8 provides an analysis of the Document Master File by bibliographic element. Included as part of the report is an exception listing of bibliographic entries not in agreement with the current list of authorized elements.

Programmer's Note

Valid bibliographic elements. MAINT8 is one of the few programs that checks element

number against a pre-programmed list rather than a lead card, that is a program control card providing data required for processing. The authorized bibliographic elements have been changed since the program was first implemented. As a result, some error messages will appear that require no action by the user. (See Chapter X, specifically the section on MAINT8.) If so desired, the list of element numbers could be revised by reassembling the program. Only minor programming modifications are required.

MAINT9 – Subject the Document Master File

Abstract

With a continually increasing file size, it is sometimes necessary to work with a portion of the file – perhaps a class of documents or specific documents. MAINT9 allows the Document Master File to be subdivided into groups based on document number. Subdivision can be as broad as document class or as specific as individual document with many alternatives in between. The subdivided file can be used in exactly the same ways as the complete Document Master File.

LIBPRT – Library Print of the Document Master File

Abstract

MAINT6 and LIBPRT are part of the same program. Both are designed to print the contents of the Document Master File; their functions vary in the amount of information printed and in format. MAINT6 provides a listing of the transactions to the Document Master File for any one update cycle while LIBPRT provides a listing of the entire Document Master File. (See *General Manual,* Chapter VI, specifically the File Maintenance Module Tasks and Figure 17.) When using LIBPRT, no provision is made for creating a Descriptor Input File.

SORTD1 – Sort the Descriptor Input File

Abstract

Since the Descriptor Input File will update the contents of the Inverted File, the two files must be in the same sequence. The Descriptor Input File is created in document number sequence and sorted to descriptor and document number sequence.

DMAIN1 – Update the Inverted File

Abstract

The Inverted File will be updated with the contents of the Sorted Descriptor Input File. Because this file is created directly from the update of the Document Master File (MAINT4, MAINT6), the integrity of the Inverted File is assured. The Inverted File will be updated with the postings for each descriptor used, including document numbers, frequency count and dates of latest activity.

DMAIN2 – Print the Summary Listing of the Inverted File

Abstract

The Summary Listing of the descriptors reports on the terms used in document analysis. This listing is useful in formulating search requests, evaluating the indexing procedures and reviewing the Thesaurus File since it provides (by term) the term type, the frequency of usage, and the last date of activity on the Document Master File.

Programmer's Notes

Internal descriptors. URBANDOC has created several internal descriptors, descriptors included for specific retrieval needs but which hold little or no value for the external user. An example of this is the issue number of the *Input Index* in descriptor form. This is also true for the variation of the document number as a descriptor; since document number is always shown with the reference there would be no value to repeat it as a series of entries, one for each reference, in the printed Inverted File.

Subsetting the Inverted File. An internal descriptor is a category of descriptor innovated by URBANDOC and used primarily for document numbers in a descriptor form. These internal descriptors are identified by a pre-code of '-' instead of a pre-code of '*'. (See Chapter VII, specifically Thesaurus Data Entry.) The purpose of creating this type of descriptor is that these terms can, and should, be eliminated from the Summary Listing. Inclusion of these terms would not increase the value of the report but rather unnecessarily lengthen it.

DMAIN3 – Print a Detail Listing of the Inverted File

Abstract

The Detail Listing is a supplemental report to the Summary Listing, identifying the document numbers of the references to which the terms have been assigned. The Major Subject Listing in the *Input Index* provides a means of manually searching for the broad subject headings assigned to a document. The Detail Listing becomes significant since it is the only subject access to all the descriptors assigned to a document.

Programmer's Notes

Expansion of the report format. The user has the option of including the data presented in the Summary Listing as part of this report. Due to file organization and method of programming, frequency count cannot exceed one hundred and fifty per term. For terms with a greater frequency of usage, all postings will be shown but frequency count will appear as one hundred and fifty. For a correct frequency count for these terms, the user refer to the Summary Listing.

Subsetting the Inverted File. The exclusion of the internal descriptors should also apply to this report. The presence of these terms would not increase the value of the report but rather unnecessarily lengthen it.

Tape File Specifications

Docu-To-Tape File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	'1HDR'
Filler	X(4)	5- 8	
Label	X(12)	9-20	'DOCU-TO-TAPE'
Filler	X(10)	21-30	
Date	X(6)	31-36	mmddyy
Reel Number	X(3)	37-40	

c. *Data Record Format:*

Fixed-length records of 84 characters
Blocking factor of 10
Padding record of 9s

Five data records:
- Header
- Descriptor
- Subdescriptor
- Bibliographic Data
- Subject Heading

d. *Header Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Filler	X(2)	21-22	
Data Entry Date	X(24)	23-46	'ENTRY DATE mm/yy
Filler	X(38)	47-84	

e. *Descriptor Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	

Filler	X(2)	21-22	
Pre-code	X	23	
Descriptor	X(35)	24-58	
First Subdescriptor	X(12)	59-70	Used only for publications date
Numerical Value	X(10)	71-80	
Filler	X(4)	81-84	

f. *Subdescriptor Format:*

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Unit Number	X(4)	15-18	
Entry Code	X	19	
Entry Type	X	20	
Descriptor Tie	X(2)	21-22	
Subdescriptor Number	X(2)	23-24	
Descriptor Root	X(6)	25-30	
Filler	X(4)	31-34	
Subdescriptor	X(12)	35-46	Used only for entry date or content dates
Numerical Value	X(10)	47-56	Used only for entry date or content dates
Filler	X(2)	57-58	
Subdescriptor	X(12)	59-70	Used only for content dates
Numerical Value	X(10)	71-80	Used only for content dates
Filler	X(4)	81-84	

g. *Bibliographic Data Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Bibliographic Text	X(58)	23-80
Filler	X(4)	81-84

h. *Subject Heading Format:*

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Element Number	X(2)	21-22
Subject Code	X	23
Subject Sequence	X	24

Subject Heading	X(35)	25-59
Filler	X(25)	60-84

i. *Trailer Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(36)	5-40	

Look-Up Input File
Sorted Look-Up Input File

a. *File Format:*
Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	'1HDR'
Filler	X(4)	5- 8	
Label	X(13)	9-21	'LOOK-UP INPUT'
Filler	X(12)	22-33	
Reel Number	X(3)	34-36	
Filler	X(4)	37-40	

c. *Data Record Format:*
Fixed-length records of 56 characters
Blocking factor of 20 records
Padding records of 9s

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Pre-code	X	21
Term	X(23)	22-44
Filler	X(12)	45-56

d. *Trailer Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(36)	5-40	

Bad Look-Up File
Sorted Bad Look-Up File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 40 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	'1HDR'
Filler	X(4)	5- 8	
Label	X(11)	9-19	'BAD LOOK-UP'
Filler	X(11)	20-30	
Date	X(6)	31-36	mmddyy
Filler	X	37	
Reel Number	X(3)	38-40	

c. *Data Record Format:* Unblocked fixed-length records of 59 characters.

Field	Size	Cols
Document Number	X(14)	1-14
Unit Number	X(4)	15-18
Entry Code	X	19
Entry Type	X	20
Pre-code	X	21
Term	X(35)	22-56
Error Code	X(3)	57-59

d. *Trailer Label Format:* Single fixed-length record of 40 characters.

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'
Filler	X(36)	5-40	

Edit File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 24 characters.

Field	Size	Cols	Contents
Label	X(15)	1-15	'EDIT-TAPE-PASS1'

Reel Number	X(3)	16-18	
Date	X(6)	19-24	mmddyy

c. *Data Record Format:* Unblocked variable-length records, maximum of 161 characters

Field	Size	Cols	Contents
Document Number	X(14)	1- 14	
Sequence Number	X(3)	15- 17	
Entry Code	X	18	
Type Code	X	19	
Revision Data	X(11)	20- 30	Used only for Document Master Revisions
Filler	X(2)	31- 32	
Input Unit Number	X(4)	33- 36	
Text Size	X(3)	37- 39	
Text	X(text size)	40-161	Variable in length, maximum of 122 characters

d. *Trailer Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(20)	5-24	

Document Master File
Document Subset File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Label	X(12)	1-12	'DOCU/MASTERS'
Filler	X(3)	13-15	
Reel Number	X(3)	16-18	
Date	X(6)	19-24	mmddyy

c. *Data Record Format:* Unblocked variable-length records, maximum length of 2200 characters. See also the discussion of the Document Master File

Two data records:
Searchable Record
Free Text Record

d. *Searchable Record Format:* Unblocked variable-length record, maximum length of 2200 characters

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Element Number	X(3)	15-17	
Data Entry Date	X(24)	18-41	'ENTRY DATE mm/yy
Filler	X(14)	42-55	
Pointer Table	X(3)+NX(11)		N = number of descriptors assigned. See the discussion of the Document Master File.
Searchable Data			Variable in length to 2200 characters.

e. *Free Text Record Format:* Unblocked variable-length records, maximum length of 2200 characters

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Element Number	X(3)	15-17	
Data Entry Date	X(24)	18-41	'ENTRY DATE mm/yy
Filler	X(14)	42-55	
Free Text Data			See the discussions of the Document Master File. Variable in length to 2200 characters.

f. *Trailer Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Identification	X(3)	1- 3	For a multi-reel file, only the last reel contains 'EOF'; all other reels contain 'EOR'.
Filler	X(21)	4-24	

Print File

a. *File Format:*

Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Label	X(10)	1-10	'PRINT TAPE'
Filler	X(5)	11-15	
Reel Number	X(3)	16-18	
Date	X(6)	19-24	mm/dd/yy

c. *Data Record Format:*

Three data records:
- Descriptor Transaction
- Searchable Record
- Free Text Record

d. *Descriptor Transaction Format:* Unblocked fixed-length records of 43 characters

Field	Size	Cols
Transaction Code	X(2)	1- 2
Pre-code	X	3
Descriptor	X(23)	4-26
Document Number	X(14)	27-40
Element Number	X(2)	41-42
Record Mark	X	43

e. *Searchable Record Format:* Unblocked variable-length record, maximum length of 2200 characters. See also discussion of the Document Master File

Field	Size	Cols	Contents
Entry Code	X(2)	1- 2	
Document Number	X(14)	3-16	
Element Number	X(3)	17-19	
Data Entry Date	X(24)	20-43	'ENTRY DATE mm/yy
Filler	X(14)	44-57	
Pointer Table	X(3)+NX(11)		N = number of descriptors. See the discussion of the Document Master File.
Searchable Data			Variable in length to 2200 characters.

f. *Free Text Record Format:* Unblocked variable-length record, maximum length of 2200 characters. See also the discussion of the Document Master File

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Element Number	X(3)	15-17	
Data Entry Date	X(24)	18-41	'ENTRY DATE mm/yy
Filler	X(14)	42-55	
Free Text Data			See the discussion of the Document Master File. Variable in length to 2200 characters.

g. *Trailer Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Identification	X(3)	1- 3	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(21)	4-24	

Descriptor Input File
Sorted Descriptor Input File

a. *File Format:*
Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 41 characters

Field	Size	Cols	Contents
Label	X(19)	1-19	'DESCRIPTOR-PASS ØØ1'
Date	X(6)	20-25	mmddyy
Filler	X(16)	26-41	

c. *Data Record Format:*
Fixed-length records of 42 characters
Blocking factor of 4
Padding records of 9s

Field	Size	Cols
Entry Code	X	1
Filler	X	2
Descriptor Type	X	3
Descriptor	X(23)	4-26
Document Number	X(14)	27-40
Filler	X(2)	41-42

d. *Trailer Label Format:* Single fixed-length record of 41 characters

Field	Size	Cols	Contents
Identification	X(4)	1- 4	For a multi-reel file, only the last reel contains '1EOF'; all other reels contain '1EOR'.
Filler	X(37)	5-41	

Inverted File

a. *File Format:*
Header label, tape mark
Data records, tape mark
Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 41 characters

Field	Size	Cols	Contents
Label	X(15)	1-15	'DESCRIPTOR-FILE'
Reel Number	X(3)	16-18	

Filler	X(2)	19-20	
Date Created	X(12)	21-32	'DATE CREATED'
Filler	X	33	
Date	X(8)	34-41	mm/dd/yy

c. *Data Record Format:* Unblocked variable-length records, maximum length of 2157 characters

Field	Size	Cols	Contents
Descriptor	X(23)	1-23	
Filler	X	24	
Descriptor Type	X	25	
Frequency Count	X(5)	26-30	
Date of Last Addition	X(6)	31-36	mmddyy
Date of Last Deletion	X(6)	37-42	mmddyy
Highest Document Number for Descriptor	X(14)	43-56	Blank for common descriptor
Record Mark	X	57	
Document Number	NX(14)	58-end	N = number of documents. Occurs maximum of 150 times only for precise or internal descriptors

d. *Trailer Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Identification	X(3)	1- 3	For a multi-reel file, only the last reel contains 'EOF'; all other reels contain 'EOR'.
Filler	X(21)	4-24	

References to the URBANDOC Final Report

Much of the information presented in this chapter is designed to be used with sections of the *General Manual* (G.M.) and other sections of the *Operations Manual* (O.M.)

For additional information on the Document Master File and the Inverted File, their formats and the considerations in creating and maintaining entries on the files, see:

Manual	Chapter	Section
G.M.	I – Introduction	The Bibliographic Records
G.M.	II – Document Identification	URBANDOC Document Numbers: General
G.M.	III – Document Analysis: Descriptive	General Considerations
G.M.	IV – Document Analysis: Content	General Considerations
G.M.	VI – Systems Modules: Input	File Maintenance Module: Document Master File Description, Inverted File Description

For additional information on the design and goals of the File Maintenance Module, see:

Manual	Chapter	Section
G.M.	VI – Systems Modules: Input	File Maintenance Module: Function, Tasks

For additional information on operating this portion of the system, see:

Manual	Chapter	Section
O.M.	VIII – Processing Cycles	Editing and Validation Cycle, Input Processing Cycle, *Retrieval Report* File Maintenance Cycle, Miscellaneous Publications Products
O.M.	IX – Operating Instructions	File Maintenance Module
O.M.	XI – Tape Library and Report Controls	File Maintenance Module
O.M.	XII – Timing	File Maintenance Module

SEARCH MODULE

Introduction

The Search Module in the URBANDOC system is a composite of two separate subsystems: the CFS Search Module and the *Enginneering Index* Search Expansion Programs (SEP). These two have been interfaced together to form a processing whole. Originally, only the CFS programs were used. The *Engineering Index* programs were subsequently added.

> Knowledge of the basic concept underlying the Combined File Search process is useful in understanding the search run. The Combined File Search uses both a Descriptor File and the Master File in the search process. This process is based upon the following principle:
>
> A search request contains a number of descriptors. If we take any one of these descriptors . . . and obtain the list of items posted under that descriptor in the Descriptor File, then we have immediately reduced the possible responses to the request from the entire Master File to that limited item list. It is then possible to go to the Master File and perform a detailed search on the selected items. Thus, the actual search is performed on the Master File; the Descriptor File is only a tool for reducing the scope of the search.[1] (See Figure 5.)
>
> The search run consists of all processing necessary to produce the output requested by the user of the system. The process begins with the processing of a card file containing the users' requests and terminates when the documents selected by the search have been printed out. This process is accomplished in six phases. (See Figure 6.)

Phase	*Function*
1	Card-to-tape and edit the request card file
2	Descriptor File search
3	Sort output from Phase 2 into item number sequence
4	Associate sorted item numbers from Phase 3 with full request terms
5	Master File Search
6	Print search run output.[2]

Experience with the released search portion of the CFS system revealed three significant limitations:

> The format of the printed search output was generally unacceptable for widespread usage, especially without annotation;
>
> The formation of complex search expressions was cumbersome to the analyst. On

[1] International Business Machines, Data Processing Division, *1401 Information Storage Retrieval System (The Combined File Search System)* by Donald Prentice, Gary deGraw, Alice Smith and I. Albert Warheit, Rev. III, 1 v. (White Plains, N. Y., 1966), p3.01.

[2] *Ibid.*, p 3.04.

occasion the resulting search expressions did not agree with the conceptualized definition of the search request;

The sequencing of criteria within the search terms was the responsibility of the document analyst. Usually the formulated search expression did not run at maximum search efficiency, i.e. the terms were not ordered with the least frequent terms first. (A joint review of the request by the document analyst and systems analyst would have resolved this problem but would have been costly in terms of man-time.)

URBANDOC became aware of a series of programs called the Search Expansion Program (SEP) designed and implemented by the *Engineering Index* specifically to overcome the above limitations of the CFS system. URBANDOC acquired these programs and interfaced them with the existing CFS system. (See Chapter I, specifically Systems Background.)

The Search Expansion Programs (SEP) were added to the CFS Search Programs to reduce the amount of time spent in writing a request, to minimize the time required to perform a search and to produce the search output in a more acceptable format. With CFS, the document analyst was required to work with a cumbersome data entry procedure. This was particularly true when one term was used more than once in a request, for example A*(B+C) which became A*B+A*C in the CFS format. The SEP programs would now perform the expansion. The search time could be minimized by placing the least used terms at the beginning of the request. More importantly, the SEP programs relieved the document analyst of the burden of manually sequencing the term. These points will be discussed in greater detail in the Narrative (next section).

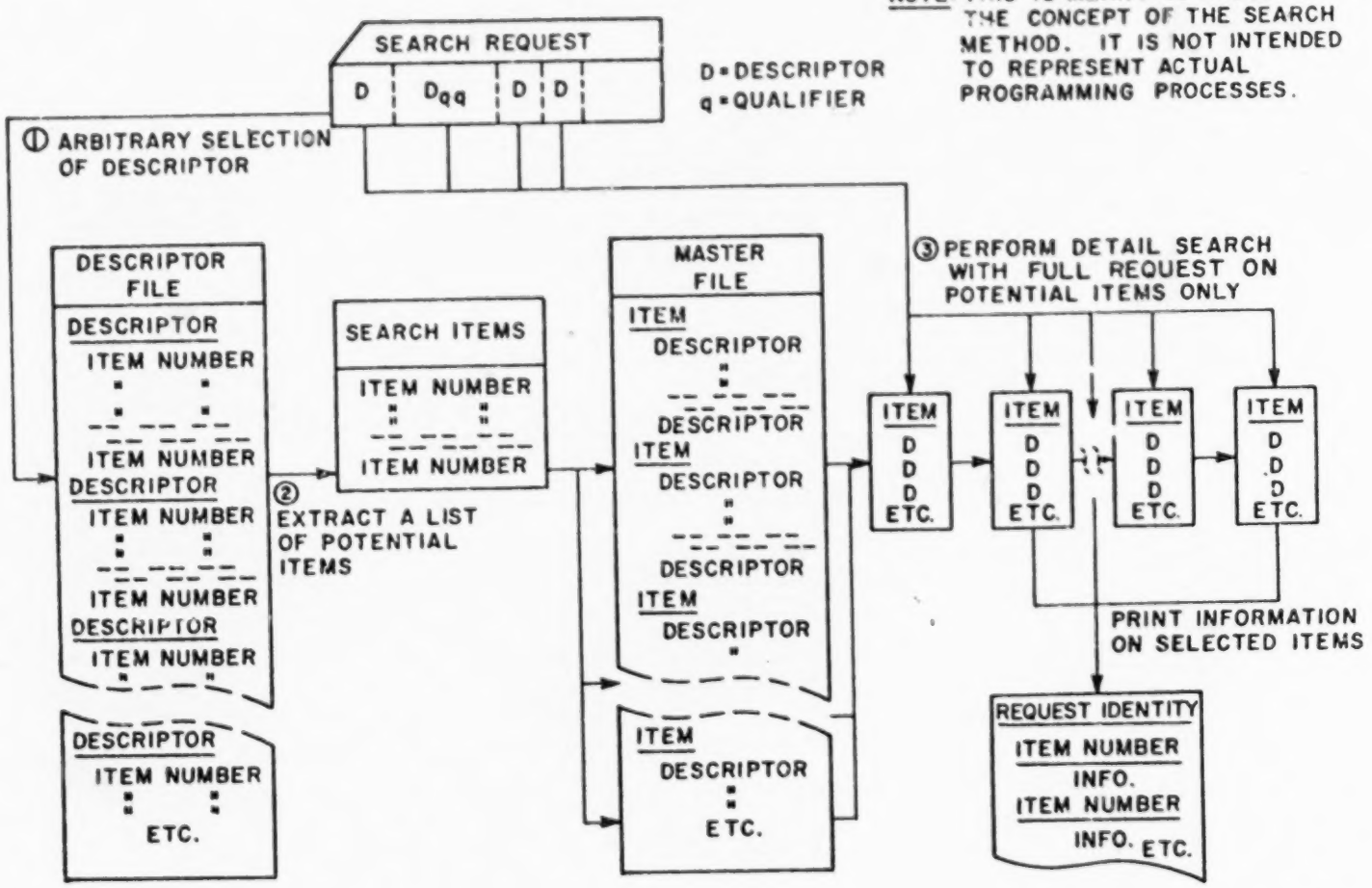
The SEP system consists of two groups of programs. One group is a set of diagnostic programs, designed to detect potential processing problems in the CFS search prior to the actual search. The other is a series of formatting and print programs designed to replace their corresponding counterparts in CFS.

The search output is the documents meeting the search requirements (the hits) listed for each request. The entry in the *Retrieval Report* will consist of the title information from the input request plus the entire record for that document in the Document Master File: descriptors, authors, title, imprint and collation, subject headings, etc. The term *"Retrieval Report"* is an URBANDOC label for the printed search output; it is not a specific part of the CFS or the *Engineering Index* systems.

In its current form, a search request may be processed either as part of the Expanded Search (including the use of the SEP system) or as a Basic Search (using only the CFS portion of the search). Depending upon the search method selected, the corresponding form for Search Data Entry will be used. (This will be discussed in greater detail in the following section on Narrative. See also Chapter VII, specifically Search Data Entry.)

The following section on Narrative provides a more detailed discussion of the actual processing techniques used in the composite search system. If the reader only wishes an overview of the steps involved in processing a search request, see Program Abstracts and omit the Narrative.

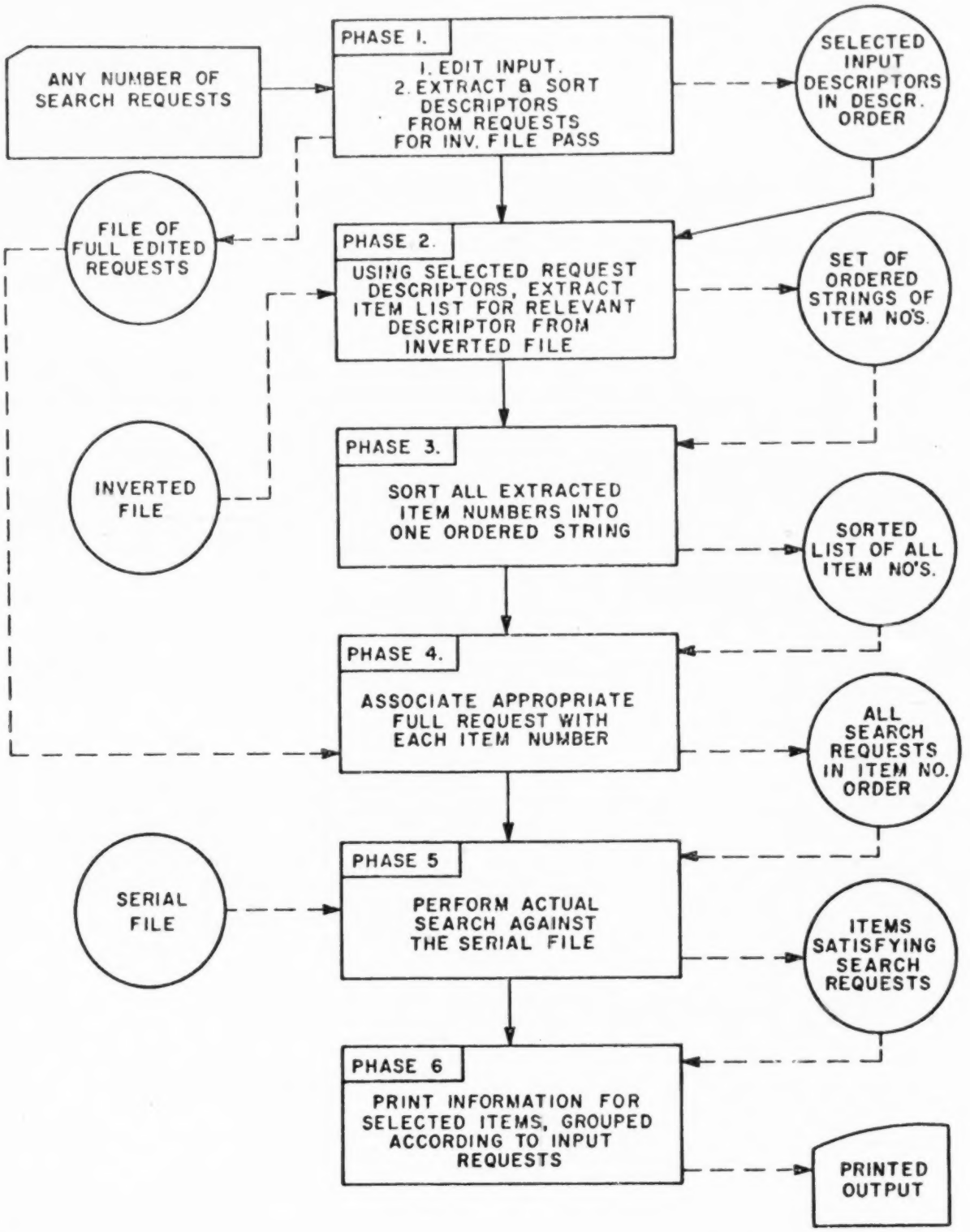
CONCEPT OF THE COMBINED FILE SEARCH METHOD[3]



[3] *Ibid.*, p 3.02.

Figure 5

LOGICAL FLOW OF COMBINED FILE SEARCH PROCESS[4]


ANY NUMBER OF SEARCH REQUESTS
PHASE 1.
1. EDIT INPUT.
2. EXTRACT & SORT DESCRIPTORS FROM REQUESTS FOR INV. FILE PASS
SELECTED INPUT DESCRIPTORS IN DESCR. ORDER
FILE OF FULL EDITED REQUESTS
PHASE 2.
USING SELECTED REQUEST DESCRIPTORS, EXTRACT ITEM LIST FOR RELEVANT DESCRIPTOR FROM INVERTED FILE
SET OF ORDERED STRINGS OF ITEM NO'S.
INVERTED FILE
PHASE 3.
SORT ALL EXTRACTED ITEM NUMBERS INTO ONE ORDERED STRING
SORTED LIST OF ALL ITEM NO'S.
PHASE 4.
ASSOCIATE APPROPRIATE FULL REQUEST WITH EACH ITEM NUMBER
ALL SEARCH REQUESTS IN ITEM NO. ORDER
SERIAL FILE
PHASE 5
PERFORM ACTUAL SEARCH AGAINST THE SERIAL FILE
ITEMS SATISFYING SEARCH REQUESTS
PHASE 6
PRINT INFORMATION FOR SELECTED ITEMS, GROUPED ACCORDING TO INPUT REQUESTS
PRINTED OUTPUT

[4] *Ibid.*, p 3.03.

Figure 6

Narrative

The *General Manual* described the search process as one total entity. However, a detailed examination of the Search Module would reveal that it is a series of subsystems, separately developed by different organizations, which have been interfaced to form a whole.

In order to discuss the Search Module to the depth that it warrants URBANDOC has inserted parts of narrative from the author organization documentation. URBANDOC comments and explanations are inserted for a smooth transition from one topic to another.

> Three types of searches may be specified by the user of this system, DOCUMENT, BOOLEAN, or MIXED.
>
> Document Search. Purpose: To retrieve the information pertaining to a particular document or series of documents. Input Data Required: List of document numbers for which information is to be retrieved.
>
> Boolean Search. Purpose: To retrieve document numbers and information related to an object which may be described as using a Boolean statement. Input Data Required: A list of criteria which describes the object worded in a Boolean statement.
>
> Mixed Search. Purpose: To retrieve information related to a specific document or series of documents provided they conform to a given set of criteria. The Mixed Search is a combination of the Document Search and the Boolean Search. Input Data Required: A list of documents to be examined and the descriptors which describe the criteria for which the documents are to be examined.[5]

URBANDOC usually performs a Boolean or Descriptor Search. For this reason, the Document Search or Mixed Search is not discussed at length in the URBANDOC Manuals. Both the Boolean Search and the Mixed Search use the Inverted File for an inverted table look-up followed by a linear search of the Document Master File. The Document Search does not perform the inverted table look-up but is a straight linear search of the Document Master File.

> Three data files are required in the search process: Request File, Descriptor File, Master File.
>
> Request File: This file is a card input batch of assembled search requests, each request containing the criteria for search and retrieval as specified by the requestor. (Note: Here the term, Request File, is the file called the Basic Search Input in the *Operations Manual.*)
>
> Descriptor File: The file is a series of items, or documents, each of which contains all information indexed for the particular item.

[5] *Ibid.*, p 3.21

Master File: The file is a series of items, or documents, each of which contains all information indexed for the particular item.[6]

The Document Master File and the Inverted File have been previously discussed in greater detail. (See Chapter IV, specifically the Document Master File and the Inverted File.)

Each request consists of three distinct sections, each of which must be present:

Title Section: To identify the individual making the request. The identifying information will appear at the top of each page of printed output.

Format Section: The format section designates the type of search being made (one of three) and the desired format of the output.

Request Section: The body of the request is represented in the section. The information in the request section will vary depending on the type of search specified.[7]

The basic mechanics of the search are as follows: *Document Search:* If the search is a 'DOCUMENT' search, the record is retrieved unconditionally. *Boolean or Mixed Search:* The record is searched for the first criteria in the query. If the criterion is satisfied, the search proceeds with the next criterion. If the end of the request is encountered before any criterion fails to be met, the record is selected. If any criterion is not met, the search will read the next request from the document tape and the process begins again.

The search program operates on the principle that every document searched represents a 'hit', unless it fails to meet the criteria specified in the query. In other words, if the criteria list becomes exhausted before the document fails to meet any of the specified criteria, the document will be retrieved.[8]

The following terms in the search narrative have specific meaning in the CFS documentation:

Criterion	A descriptor with or without a qualifier.
Descriptor	That part of a criterion which describes a property of a document
Qualifier	A term which narrows the meaning of a descriptor
Subdescriptor	The alphabetic portion of a qualifier[9]

[6] *Ibid.*, pp 3.01, 3.04.

[7] *Ibid.*, p 3.18.

[8] *Ibid.*, p 3.15.

[9] *Ibid.*, p 3.18

Three Boolean Operators are used by the system to define the logical relationship between the descriptors within a request. These operators take on a precise meaning.

Except for the Document Search the criteria are listed in a Boolean Expression. The criteria within the Boolean Expression are connected by Boolean Operators.

Operator	Meaning	Comments
*	*AND*	A*B, where A and B represent descriptors, indicates that we are searching for a document which has the property A *and* the property B.
/	*AND NOT*	A / B indicates that we are searching for a document which possesses the property A, but does *not* possess the property B.
+	*OR*	A + B means we are seeking a document possessing property A, *or* property B *or* both properties[10].

Except when preceded by the *NOT* operator '/', the inclusion of a descriptor in a search request indicates that the document references will be examined for the presence of that descriptor. Once the descriptor's presence has been ascertained, further qualifications or narrowing on the descriptor may be done, either for the presence or absence of specific natural language subdescriptors or to search on numeric values.

Two operators are used to show the relationship between the descriptor and subdescriptor. *'.EQ.'* shows that the presence of the subdescriptor is required. *'.NE.'* shows that the presence of the subdescriptor is prohibited.

Relations between descriptors and values or subdescriptors and values are governed by arithmetic operators. In the actual request, the operator precedes the entire qualifier even though it applies to the value. The relationship between the descriptor and subdescriptor is always an implicit *'AND'*. The operators follows:

Arithmetic Notation	*Meaning*
.NE.	Not equal
.GE.	Greater than or equal to
.GT.	Greater than
.EQ.	Equal to
.LT.	Less than
.LE.	Less than or equal to[11]

[10] *Ibid.*, p3.22

[11] *Id.*

These symbols could also be considered as relational operators or comparative operators. The term arithmetic operator seems to suggest relationships that are strictly numerical in nature. This is true for the relationship between the descriptor or subdescriptor and a value; it is not true of the relationship between the descriptor and subdescriptor. When the above symbols are used with numerical values, they could be considered true arithmetic operators since they are being used in a completely numerical capacity.

URBANDOC uses numerical qualifiers only for date analysis. For this reason, it is not discussed in detail in the URBANDOC manuals. However, for an example, of how this feature would be used:

DATES .EQ. PUBLICATION 1968 would be a valid request where DATES is a descriptor and PUBLICATIONS is a qualifier. In this case, the value of PUBLICATIONS would have to be 1968 for selection to occur. "The arithmetic qualifier specifies the desired relationship of the qualifier to the descriptor. When a qualifier contains both the subdescriptor and a value, the subdescriptor must precede the value. If a descriptor required more than one qualifier, the descriptor must be repeated."

The criteria within the Boolean Expression are grouped together as Search Terms:

> A Search Term is defined as a segment separated by *OR*. The request A * B + C * D (meaning A and B or C and D) may be thought of as two requests – Request A * B and request C * D, since either A * B or C * D will result in selection of a document independently of the other. In the request A * B + C * D, A * B and C * D are both separate terms. [A *AND* B) *OR* (C *AND* D) will be satisfied by the presence of A and B, C and D or both A, B, C and D.]
>
> Every request must begin with a *Precise Descriptor* which may *not* have an operation sign preceding it. The first descriptor of a term, as all descriptors, may or may not be qualified at the user's will.
>
> The question now arises as to how the user of the system will know which descriptors are Precise Descriptors and which ones are Common Descriptors. A . . . Descriptor list will be produced as part of the Master File maintenance function which will list each descriptor, indicate whether the descriptor is precise or common, and specify how many documents on the Master File contain the descriptor[12] .

The CFS system established two types of descriptors: "precise descriptors" and "common descriptors". A "common descriptor" was defined by CFS to be those terms with an extremely high frequency of usage. All terms not defined as "common" were termed "precise". A "common descriptor" may not be used as the only term in a search. "Precise descriptors" are said to have a stand-alone value in searching, that is they may be used as the only term in a search. URBANDOC has defined such terms as DATES and ARTICLES as "common descriptors".

[12] *Ibid.*, p 3.23.

If the requestor has a list of criteria which, when written out, contains several 'AND'-ed criteria per term(s), he will realize a saving in processing time by placing the descriptor with the least number of occurrences at the beginning of the term(s).

Assume we have the following entries in the . . . Descriptor List:

Descriptor	*Type*	*Occurrences*
A	Precise	155
B	Precise	623
C	Precise	14
D	Common	497
E	Precise	650
F	Precise	97

If we wish to retrieve a document with the property A and the property C and not the property E, or with the property B and not the property D and not the property F, we could phrase the request several ways:

(A * C / E + B / D / F)
Term 1 Term 2

Note: All Boolean statements must be enclosed by a single set of parentheses.

Term 1 may also be written as C * A / E. This form will result in a substantial savings in processing time. Primary document selection is dependent on the first descriptor in each term. If Term 1 were written A * C / E, the system would be required to closely examine 155 documents for Term 1, whereas, the system would only closely examine 14 documents if Term 1 was rephrased C * A / E. Now examine Term 2. Since D is a common descriptor we may not use it as the first descriptor in Term 2. Also we may not use descriptor F as our first descriptor in Term 2 because of the *AND NOT* operation. Thus, Term 2 must stand as it is and the system will be required to closely examine 623 documents. The user may not always be free to rephrase his request in a time saving manner, but when possible, it should be considered.[13]

There are two other features which may be used during a search, truncation and sequential array.

When a requestor uses a '$' as the last character of a descriptor or subdescriptor (to indicate truncation) the unit type character in the logic control block signals the search program to shorten the Master File record term to the same number of characters as the request term before making a comparison. (See *General Manual,* Chapter V, specifically Broadening the Search, Truncation, and Search Statements. URBANDOC uses this feature for searching on geographic descriptors.)

The sequential array is signaled by a '1' as the sequential array character of the logic control block. The sequential array process will scan the record both forward and

[13] *Ibid.,* pp 3.23-3.24.

backward to determine whether data is present in a given sequence. Special indexing considerations must be made to make this feature of the search possible.[14] (This feature is not used by URBANDOC. For this reason, it will not be discussed in the URBANDOC manuals.)

Once the search has been performed, the output processing occurs in two stages. The first step is to convert the request from the search form back to the Boolean Expression. The second step is the printing of the retrieval results (*Retrieval Reports*).

All output is separated by requestor to provide for efficient routing. The page heading information is printed followed by the original request – as interpreted by the program. The system allows the individual to specify which parts of the retrieved information is to be printed:

Document number and title – always prints, followed by one of the following options:

1. Searchable section
2. All free text
3. Selected portions of free text
4. Searchable section and selected portions of free text

Numbers may be printed in two formats:

1. Normal decimal format (123.456)
2. Engineering notation (.123456x 03)[15]

Since URBANDOC uses numerical information only for date analysis, normal decimal format is always the option for the URBANDOC system. A generalized form for the search output as designed under the CFS system is illustrated in Figure 7.

The following excerpts from the *Engineering Index* documentation reveal the benefits gained from the SEP system:

The programs utilized a form of mathematics known as Polish Notation, which would expand a Boolean statement.

Ex. (A+B+C)*(D+E+F)

A*D+A*E+A*F+B*D+B*E+B*F+C*D+C*E+C*F

This System reduces the number of times a term must be entered in a request. The system also assures that a request that has passed through the search expansion programs should not be rejected by PHASE1 of the CFS Search Program.

The Indexer fills out the worksheet, on which he will assign A-Z, 1-9 to the terms

[14] *Ibid.*, p 3.28.
[15] *Ibid.*, p 3.16.

CFS OUTPUT AND FORMAT SPECIFICATIONS[16]

Request Page No.

First entry on every print page of request output

1. *Request Identification Information*
   Reproduced directly from input (maximum of 2 print lines of information). Typical information would be:

   Requester name, requester dept., etc.

On first page only

2. *Full Request* (In format as altered for search process)

3. *Record Number and Record Title*
   Precedes every record reported

4. *Searchable Section* (Only if requested on input)

   This section printed in entirety only. Print specifications are as follows:

   (sp = spaces) (underlined material is what is actually printed).

(25 sp) *Searchable Section*

(10 sp) *Informative Field*

**Descriptor 1* **Descriptor 2* **Descriptor 3* etc.

1 sp 4 sp s u 1

**Descriptor* n *Subdescriptor 1* *Value Subdes–*

2 sp 1 sp 1 sp 1 sp 4 sp

s u 1

Subdescriptor n *Value* / **Descriptor n+1*

sp 4 sp

(Note: Descriptors and qualifiers must be entierly on one line).

5. *Free Text Segments* (Only if requested, up to 5 individual segments or the entire free text section may be requested).

   Printout specifications:

(25 sp) *Free Text Segment x*
Actual Free Text Information Exactly as it is Stored.
etc."

65

[16] Ibid., p. 3.29.

Figure 7

he will use in the request. He, therefore, has a maximum of thirty-five different terms that he may use when he formulates his equation. Requests may be coupled together if the indexer so desires, thereby giving him a greater maximum than thirty-five terms. These requests will be combined by the final Search Expansion Program.

The programs are:

SEP 1: This program will take the Boolean statement and expand it. In the previous example the actual expanded version would be:

A*DA*EA*FB*DB*EB*FC*DC*EC*F.

The *OR* operator '+' was dropped. This version is then written on a scratch tape while the table the indexer assigned to each term is written on another scratch tape (Table File).

S SEP 1: The Table File is then sorted in alpha sequence by term.

SEP 2: The frequency of each term is then affixed to that term from the Descriptor File.

S SEP 2: The Table File with Frequencies is then sorted in sequence by request numbers and frequency.

SEP 3: Will match the Request File against the (Sorted) Table File with Frequencies and punch out the full request which will be used as input to PHASE1 of the CFS Search.[17]

The search output are the documents meeting the search requirements (the hits) listed for each request. The entry in the *Retrieval Report* will consist of the title information from the input request plus the entire record for that document in the Document Master File: descriptors, authors, title, imprint and collation, subject headings, etc.

In its current form, a search request may be processed either as part of an Expanded Search (including the use of the SEP system) or as a Basic Search (using only the CFS portion of the search). Depending upon the search method selected, the corresponding form for Search Data Entry will be used. (See Chapter VII, specifically Search Data Entry.) It will be necessary to use the Basic Search for the following conditions:

(1) a Document Search

(2) a Mixed Search

(3) a search including the specification of date criteria, either for publication, content or entry into the system.

[17] Engineering Index, Inc., *1968 User Participation Program Level II (Tape Service) System Documentation and User's Manual for File Search,* 1 v. (New York, 1967), ppB6-B7.

Program Inventory

Program Inventory	Program Function	Source	Basic or Expanded*	Programming Language	Configuration Set**
SEP1	To expand the search request	*Engineering Index*	Expanded	Tape AUTOCODER	1
SEP2	To validate the descriptors	*Engineering Index*	Expanded	Tape AUTOCODER	1
SEP3	To sequence the descriptors by their frequency of usage with the least used terms first and format the request to CFS specifications	*Engineering Index*	Expanded	Tape AUTOCODER	1
PHASE1	To perform the CFS edit of the requests	CFS	Basic or Expanded	Tape AUTOCODER	Either 1 or 2
PHASE 2	To perform the inverted file search	CFS	Basic or Expanded	Tape AUTOCODER	Either 1 or 2
PHASE 3	To sort the document list from primary selection into document number sequence	CFS	Basic or Expanded	Tape AUTOCODER	Either 1 or 2
PHASE4	To associate the request terms with the CFS primary selected document numbers	CFS	Basic or Expanded	Tape AUTOCODER	Either 1 or 2
PHASE5	To perform the serial file search	CFS	Basic or Expanded	Tape AUTOCODER	Either 1 or 2

Program Inventory	Program Function	Source	Basic or Expanded*	Programming Language1	Configuration Set**
PHASE6 PASS1 (Basic)	To print output from search	CFS	Basic	Tape AUTOCODER	Either 1 or 2
PHASE6 PASS1 (Expanded)	To print output from search	*Engineering Index*	Expanded	Tape AUTOCODER	1
PHASE6 PASS2 (Basic)	To print output from search	CFS	Basic	Tape AUTOCODER	Either 1 or 2
PHASE6 PASS2 (Expanded)	To print output from search	*Engineering Index*	Expanded	Tape AUTOCODER	1

**Basic: CFS Search.
Expanded: includes expanded input form and output format of *Retrieval Report.*

**Set 1: 12K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; 1311 disk drive; advanced programming; high-low-equal compare; sense switches.

Set 2: 9K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; advanced programming; high-low-equal compare; sense swithces.

Program Abstracts

SEP1 – Expand the Search Input

Abstract

The search input will be expanded to CFS specifications according to the rules of algebra and Polish String Notation. The expanded requests will be edited for sequence errors, clerical errors within the request and violations of the CFS limitation on the size of the request. An edit list of detected errors will be produced. The individual descriptors from the Term Table Section of the valid requests will be included as part of the Table File. Since subdescriptors are not entered on the Inverted File, they will not be processed by the SEP programs. The expanded form of the requests will also be printed.

SORT SEP1 – Sort the Table File

Abstract

The Table File is created in request number and term table entry sequence. In order to match the individual descriptors from the term table against the Inverted File for usage in SEP2, the file is sorted to term sequence.

SEP2 – Validate the Table File

Abstract

The entries in the Term Table Section of each request will be matched against the Inverted File. In order to fulfill the request, the terms expressing the request must be in the Inverted File. Unused terms will be flagged for the document analyst's attention. The frequency of usage will be appended to the validated terms.

SORT SEP2 – Sort the Table File with Frequencies

Abstract

The terms from the requests have been checked for usage on the Inverted File and frequency of usage is made part of the record. Processing efficiency during search can be assured by sequencing the terms within a request by frequency of usage. At the conclusion of the sort, the terms will be in sequence by request number and frequency of usage.

SEP3 – Format the Requests to CFS Specifications

Abstract

The Request File will be matched to the Sorted Table File with Frequencies to create the search requests according to CFS specifications. The Title and Format Sections will be repeated as entered. The Logical Expression and Term Table Sections will be combined to

create the CFS Request Section in which the descriptors and subdescriptors will be inserted in pla of the symbols. Within each string of descriptors connected in an AND relationship, the terms will be sequenced with the least used terms first, according to frequency of usage. An error listing will be produced for terms not used in the request or for symbols without a corresponding descriptor. The formatted request may be produced as either card or tape output.

Programmer's Note

Tape output versus card output. SEP3 offers the choice of search output on either cards or magnetic tape. SEP3 was designed to produce tape output; card output was included as an accomodation. The use of card output will affect the *Retrieval Report. (Retrieval Report* is the URBANDOC name for the printed search output.) The retrieved references will be valid, but the flags for descriptors used in the request will not be correct. In some cases the flags will be missing, in other cases they will not correspond to the term as completed in data entry.

PHASE1 – To Perform the CFS Edit of the Requests

Abstract

PHASE1 will edit the requests according to the specifications of the CFS system. The requests will be checked for completeness, validity of search type, and clerical errors within the body of the request. A request containing any of these errors will be listed and removed from further processing. The first two descriptors of each *AND* string will be extracted and sequenced by term for the inverted file search. The CFS Request File of the full request (used in the serial file search) and the Format Section (used to control the form of the *Retrieval Report*) is also created.

Programmer's Notes

Mode of operation. Although documented as one program, PHASE1 is four programs processed as one. Since the CFS documentation treats PHASE1 as one program from both a programming and operations viewpoint, URBANDOC has continued to do the same.

Completeness of the request. Each request must complete the Title, Format and Request Sections in that sequence. If a request is missing in any of these units or they are out of sequence, an error message will identify the cause of the error.

Clerical errors within a request. Clerical errors within a request may be exemplified by such things as improper use of blank spaces and parentheses or oversize descriptors or subdescriptors.

Use of the AND string. An *AND* string is defined as those descriptors which are logically connected by the *AND* operator for a request. Only the first two terms of each *AND* string will be separated for inclusion in the inverted file search. The decision to limit the inverted file search to two terms was made by the CFS systems team. (The

reduction-merge techniques used in PHASE3 were developed because of the memory size and input-output devices, four tape drives.) It is important to note that contrary to the CFS documentation, the first two terms (not just the first term) of each *AND* string are included.

PHASE2 – Inverted File Search

Abstract

Using the descriptors extracted from each *AND* string during PHASE1, the Inverted File will be searched for the document numbers in which each term has been used. The output of the inverted file search will be a series of document numbers (the Primary Selection File) which contain at least one of the terms in the request. In this way, the number of documents to be inspected during the serial file search may be limited to those documents with some potential of satisfying the request requirements.

PHASE3 – Sort and Reduce the Primary Selection File

Abstract

The Primary Selection File provides the key to the references on the Document Master File to be serially searched. Before this search can be performed, the lists of potential hits must be arranged in document number sequence. (The term "arrange" is used in place of "sort" since a generalized sort program is not used and other significant processing is performed.) As mentioned before, the first two descriptors of a request string were included as part of the inverted file search. During the arrangement process, any document numbers not common to both terms of a request string will be eliminated from the Primary Selection File. At the conclusion of PHASE3, the file will be in sequence by document number, request number and string number and will contain only those document numbers satisfying at least two of the search criteria.

PHASE4 – To Match the Full Requests with the Primary Selection File

Abstract

The inverted file search within the CFS system operates only on the first two descriptors of each *AND* string. It does not include the remaining terms of the string, the subdescriptors or the dates. Before performing a serial file search on the full request, the document numbers in the Primary Selection File must be associated with the appropriate requests. The output of PHASE4 is a record of the full request for which each document reference is to be interrogated.

Programmer's Notes

Associating request number and document number. The document numbers are read one at a time. The CFS Request File is then searched until the request(s) associated with the document is located and an Inquiry File record created for each. The input tapes cannot be merged since they are not processed in the manner amenable to the normal

match-merge process. It is more efficient to scan the request numbers since they will generally be outnumbered by the document numbers.

Mode of operation. Although documented as one program in the CFS documentation, PHASE4 is actually two programs: PASS1 is used for those systems with tape only; PASS2 is used for those systems with the optional 1311 disk drive. PHASE4 is not a two-pass program; one of the two passes will be selected according to computer configuration.

The use of the 1311 disk drive. If the optional disk drive is present, the Reduced Primary Selection File is written on disk to facilitate processing. The method of associating document numbers and request numbers is more amenable to disk processing.

PHASE5 – Serial File Search

Abstract

The initial processing of a search request located those documents that satisfy at least two of the criteria specified in the request. PHASE5 determines if these candidates meet the total criteria of the search: all descriptors, subdescriptors, and dates (if so specified). The document references selected as final "hits" will be included as part of the Selection File.

Programmer's Notes

Determination of a "hit". A document reference is classified as a "hit" if during the serial file search the list of criteria is exhausted before the reference fails to meet one of the search parameters.

Serial file search. The full request and document number is matched against the individual references. When a reference is a "hit", the entire reference is written on the Selection File with a "hit key" record indicating the appropriate *Retrieval Report* in which it is to appear. For a reference satisfying multiple requests, the reference is repeated with appropriate "hit key" records.

PHASE6 – To Print the Retrieval Report

Abstract

The *Retrieval Report* for each input request is printed according to the format information in the request. PHASE6 is a two-part operation: the conversion of the request from its internal form (as stored for the processing of the search) to its full form (as entered in the request); the printing of each *Retrieval Report* with page heading information, the original request and the references satisfying the request.

Programmer's Notes

Mode of operation. PHASE6 is actually two programs documented as one but operated as two separate job steps. URBANDOC has continued documenting this program as defined

in the CFS documentation. PASS1 converts the request back to its full form; PASS2 prints the *Retrieval Reports.*

Basic versus expanded mode of operation. The general description of PHASE6 covers both the basic or the expanded programs. It is the format of the *Retrieval Report* that is affected by the choice of programs. The search must, however, be completed in the mode in which it is initiated.

Assignment of search output. Although CFS provides for only printed output from a search, the Selection File could be treated as an assignment of the output to magnetic tape. Customized programming could be done to produce special purpose products from the search results, such as cumulative issues or special bibliographies.

Input Specifications

Basic Search Input
(See Chapter VII, specifically the Basic Search Worksheet)

a. *First Title Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Requestor Identification	X(4)	6- 9	
Filler	X	10	
Search Scope	X	11	
Number in Requestor Set	X(5)	12-16	
Filler	X	17	
Search Number for Output	X(2)	18-19	
Filler	X	20	
Search Date	X(4)	21-24	yymm
Filler	X	25	
Requestor Text	X(55)	26-80	

b. *Second Title Unit Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Continuation of Requestor Text	X(75)	6-80

c. *Format Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	

Unit Number	X(2)	3- 4	
Unit Code	X	5	
Field Start	X	6	'/'
Search Type	X	7	
Field Separator	X	8	','
Field Start	X	9	'/'
Descriptor Option	X	10	
Field Separator	X	11	','
Field Start	X	12	'/'
Numerical Format	X	13	
Field Separator	X	14	','
Field Start	X	15	
Text Option	X	16	
Filler	X(64)	17-80	

d. *Request Unit Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Logical Expression	X(75)	6-80

Expanded Search Input

(See Chapter VII, specifically the Expanded Search Worksheet)

a. *First Title Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Requestor Identification	X(4)	6- 9	
Filler	X	10	
Search Scope	X	11	
Number in Requestor Set	X(5)	12-16	
Filler	X	17	
Search Number for Output	X(2)	18-19	
Filler	X	20	
Search Date	X(4)	21-24	yymm
Filler	X	25	
Requestor Text	X(55)	26-80	

b. *Second Title Unit Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4

Unit Code	X	5
Continuation of Requestor Text	X(75)	6-80

c. *Format Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Field Start	X	6	'/'
Search Type	X	7	
Field Separator	X	8	','
Field Start	X	9	'/'
Descriptor Option	X	10	
Field Separator	X	11	','
Field Start	X	12	'/'
Numerical Format	X	13	
Field Separator	X	14	','
Field Start	X	15	
Text Option	X	16	
Filler	X(64)	17-80	

d. *Term Table Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Term Sequence	X	5
Descriptor	X(23)	6-28
Filler	X(13)	29-41
Subdescriptor Link	X	42
Subdescriptor	X(12)	43-54
Filler	X(26)	55-80

e. *Logical Expression Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Logical Expression	X(75)	6-80

Tape File Specifications

Table File
Sorted Table File

a. *File Format*
No header label

Data records, tape mark
No trailer label

b. *Data Record Format:* Unblocked fixed-length records of 90 characters

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Descriptor	X(23)	6-28
Filler	X(13)	29-41
Subdescriptor Link	X	42
Subdescriptor	X(12)	43-54
Filler	X(35)	55-89
Record Mark	X	90

Table File with Frequencies
Sorted Table File with Frequencies

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:* Unblocked fixed-length records of 90 characters

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Descriptor	X(23)	6-28
Filler	X(13)	29-41
Subdescriptor Link	X	42
Subdescriptor	X(12)	43-54
Filler	X(16)	55-70
Frequency	X(5)	71-75
Filler	X(14)	76-89
Record Mark	X	90

CFS Request Input File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:*

Fixed-length records of 84 characters
Blocking factor of 10
Padding record of 9s

Four data records:
- First Title Unit
- Second Title Unit
- Format Unit
- Request Unit

c. *First Title Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Requestor Identification	X(4)	6- 9	
Filler	X	10	
Search Scope	X	11	
Number in Requestor Set	X(5)	12-16	
Filler	X	17	
Search Number for Output	X(2)	18-19	
Filler	X	20	
Search Date	X(4)	21-24	yymm
Filler	X	25	
Requestor Text	X(55)	26-80	
Filler	X(3)	81-83	
Record Mark	X	84	

d. *Second Title Unit Format:*

Field	Size	Cols
Request Number	X(2)	1- 2
Unit Number	X(2)	3- 4
Unit Code	X	5
Continuation of Requestor Text	X(75)	6-80
Filler	X(3)	81-83
Record Mark	X	84

e. *Format Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Field Start	X	6	''

Search Type	X	7	'/'
Field Separator	X	8	','
Field Start	X	9	','
Descriptor Option	X	10	/
Field Separator	X	11	','
Field Start	X	12	','
Numerical Format	X	13	/
Field Separator	X	14	','
Field Start	X	15	','
Text Option	X	16	/
Filler	X(67)	17-83	
Record Mark	X	84	

f. *Request Unit Format:*

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Unit Number	X(2)	3- 4	
Unit Code	X	5	
Logical Expression	X(75)	6-80	See Chapter VII, Basic Search Worksheet
Filler	X(3)	81-83	
Record Mark	X	84	

CFS Request File

a. *File Format:*

Header label, tape mark
Request term file, tape mark
Criteria file, tape mark
Edit file, tape mark
No trailer label

b. *Header Label Format:* Single fixed-length record of 80 characters

Field	Size	Cols	Contents
Label	X(12)	1-12	'REQUEST-TAPE'
Filler	X(68)	13-80	

c. *Request Term Record Format:* Unblocked variable-length records with a maximum length of 750 characters

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
Term Number	X(2)	3- 4	
Unit Type	X	5	Unit Type through Filler occurs
Modifier Type	X	6	a variable number of times to a
Descriptor Size	X	7	maximum record size of 750

Sequential Array	X	8	characters
Link	X	9	
Link Number	X	10	
Descriptor	X(12)		Depends upon Descriptor Size.
	or X(24)		
Subdescriptor	X(12)		
Filler	X(8)		

d. *Criteria Record Format:* Unblocked fixed-length records of 32 characters

Field	Size	Cols
Unit Type	X	1
Modifier Type	X	2
Descriptor Size	X	3
Descriptor	X(24)	4-27
Request Number	X(2)	28-29
Term Number	X(2)	30-31
Record Mark	X	32

e. *Edit File Record Format:* Unblocked fixed-length records of 191 characters

Field	Size	Cols	Contents
Request Number	X(2)	1- 2	
First Title Unit	X(75)	3- 77	
Record Mark	X	78	
Second Title Unit	X(75)	79-153	
Record Mark	X	154	
Search Type	X	155	
Searchable Option	X	156	
Number Format	X	157	
Text Output	X	158	
All After	X	159	
Filler	X(2)	160-161	
Segment Numbers	X(30)	162-191	Occurs as 10 multiples of X(3)

Primary Selection File

a. *File Format:*

Header label, tape mark
Data records, tape mark
No trailer label

b. *Header Label Format:* Single fixed-length record of 80 characters

Field	Size	Cols	Contents
Label	X(8)	1- 8	'DOC-LIST'
Filler	X(72)	9-80	

c. *Data Record Format:*

Fixed-length records of 18 characters

Blocking factor of 20
Padding character of 9s

Field	Size	Cols
Document Number	X(14)	1-14
Request Number	X(2)	15-16
Term Number	X(2)	17-18

Reduced Primary Selection File

a. *File Format:*
Header label, tape mark
Data records, tape mark
No trailer label

b. *Header Label Format:* Single fixed-length record of 80 characters

Field	Size	Cols	Contents
Label	X(15)	1-15	'SORTED-DOC-LIST'
Filler	X(65)	16-80	

c. *Data Record Format:*
Fixed-length records of 18 characters
Blocking factor of 100
Padding character of 9s

Field	Size	Cols
Document Number	X(14)	1-14
Request Number	X(2)	15-16
Term Number	X(2)	17-18

Inquiry File

a. *File Format:*
Header label, tape mark
Data records, tape mark
No trailer label

b. *Header Label Format:* Single fixed-length record of 21 characters

Field	Size	Cols	Contents
Label	X(13)	1-13	'DOCUMENT-TAPE'
Filler	X(8)	14-21	

c. *Data Record Format:* Unblocked variable-length records to a maximum length of 765 characters

Field	Size	Cols	Contents
Document Number	X(14)	1-14	

Request Number	X(2)	15-16	
Term Number	X(2)	17-18	
Unit Type	X	19	Unit Type through Filler occurs a variable number of times, to a maximum record size of 765 characters.
Modifier Type	X	20	
Descriptor Size	X	21	
Sequential Array	X	22	
Link	X	23	
Link Number	X	24	
Descriptor	X(12) or X(24)		
Subdescriptor	X(12)		
Filler	X(8)		

Selection File

a. *File Format:*
   Header label, tape mark
   Data records, tape mark
   Trailer label, tape mark

b. *Header Label Format:* Single fixed-length record of 80 characters

Field	Size	Cols	Contents
Label	X(14)	1-14	'SELECTION-TAPE'
Filler	X(66)	15-80	

c. *Data Record Format:*
   Two data records:
      Hit Key
      Searchable Record
      Free Text Record

d. *Hit Key Format:* Unblocked fixed-length records of 19 characters

Field	Size	Cols	Contents
Code	X	1	'$'
Request Number	X(2)	2- 3	
Filler	X(2)	4- 5	
Document Number	X(14)	6-19	

e. *Searchable Record Format:* Unblocked variable-length record, maximum length of 2200 characters. See also discussion of the Document Master File

Field	Size	Cols	Contents
Transaction Code	X(2)	1- 2	
Document Number	X(14)	3-16	
Element Number	X(3)	17-19	
Data Entry Date	X(24)	20-43	'ENTRY DATE mm/yy

Filler	X(14)	44-57	
Pointer Table	X(3)+NX(11)		N = number of descriptors.
Searchable Data			See discussion of the Document Master File

f. *Free Text Record Format:* Unblocked variable-length record, maximum length of 2200 characters. See also the discussion of the Document Master File.

Field	Size	Cols	Contents
Document Number	X(14)	1-14	
Element Number	X(3)	15-17	
Data Entry Date	X(24)	18-41	'ENTRY DATE mm/yy
Filler	X(14)	42-55	
Free Text Data			See the discussion of the Document Master File

g. *Trailer Label Format:* Single fixed-length record of 24 characters

Field	Size	Cols	Contents
Identification	X(3)	1- 3	For a multi-reel file, only the last reel contains 'EOF'; all other reels contain 'EOR'.
Filler	X(21)	4-24	

References to the URBANDOC Final Report

Much of the information presented in this chapter is designed to be used with sections of the *General Manual* (G.M.) and other sections of the *Operations Manual* (O.M.)

For additional information on preparing requests for searches, see:

Manual	Chapter	Section
G.M.	V – Searching the Data Base	Query Analysis Techniques, Narrowing the Search, Broadening the Search, Mixed Strategies, Search Statements

For additional information on the design and goals of the Search Module, see:

Manual	Chapter	Section
G.M.	V – Searching the Data Base	Manual Versus Machine Searching
G.M.	VII – Systems Modules: Output	Search Module Function, Tasks

For additional information on operating this portion of the system, see:

Manual	Chapter	Section
O.M.	VIII – Processing Cycles	*Retrieval Report* Production Cycle
O.M.	IX – Operating Instructions	Search Module
O.M.	XI – Tape Library and Report Controls	Search Module
O.M.	XII – Timing	Search Module

PUBLICATIONS MODULE

The Publication File

The Publication File consists of a series of records for each entry in the *Input Index*. Whereas the Document Master File contains one record for each element, the Publication File contains one record for each index entry. (With the case of two joint personal authors there would be one record containing both names on the Document Master File and a record for each name on the Publication File.) Although created and maintained in sequence by document number and within document number by element number and index entry, the Publication File is sorted to other sequences for the index entry. The Publication File is sorted to other sequences for the index sections or other special reports.

The individual Publication File record contains a series of fields: document number, element number, sort control field, number of text lines and index entry. The element number is the code corresponding to the type of bibliographic information. (See Chapter IV, specifically the Document Master File.) Sort control field is a variation of the index entry for sorting purposes. (See Programmer's Notes for P0010.) The actual index entry, in its text form, may consist of a variable number of lines to a maximum of four. The number of text lines field provides the system with the size of the index entry.

Program Inventory

Program Number	Program Function	Source	Basic or On-demand*	Programming Language	Configuration Set**
P0010	To create the Publication File and the Content Analysis File	URBANDOC	Basic	COBOL	1
P0040	To create a subset of the Publication File in which there will be a citation of main entry and title for each subject heading assigned to a document	URBANDOC	Basic	COBOL	1
P0050	To print either the Main Document Listing or the Major Subject Listing	URBANDOC	Basic	COBOL	1

*Basic: Part of the standard processing of the Document Input File.
On-demand: Performed only upon request.

**Set 1: 12K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; 1311 disk drive; advanced programming; high-low-equal compare; sense switches.

Set 2: 8K memory; 1402 card reader-punch; 1403 printer; 4 tape drives; advanced programming; high-low-equal compare; sense switches.

P0060	To format and print the indexes to the Main Document Listing or produce a tape of the formatted page	URBANDOC	Basic	COBOL	1
P0070	To print formatted index pages	URBANDOC	Basic	COBOL	1
P0080	To create a special Publication File from the Document Master File or from Search Module	URBANDOC	On-demand	COBOL	1

Program Abstracts

P0010 – Create the Publication File

Abstract

The Document Input File expanded, edited and revised, updates the Document Master File and creates a Publication File for use in an *Input Index*. The Document Input File and errors detected during the Input Processing cycle are entered to ensure that the Publication File is in phase with the Edit File. With these files synchronized, the products of the *Input Index* and the Document Master File are also in phase. (See Chapter VIII, specifically Input Processing Cycle and *Input Index* File Maintenance Cycle.)

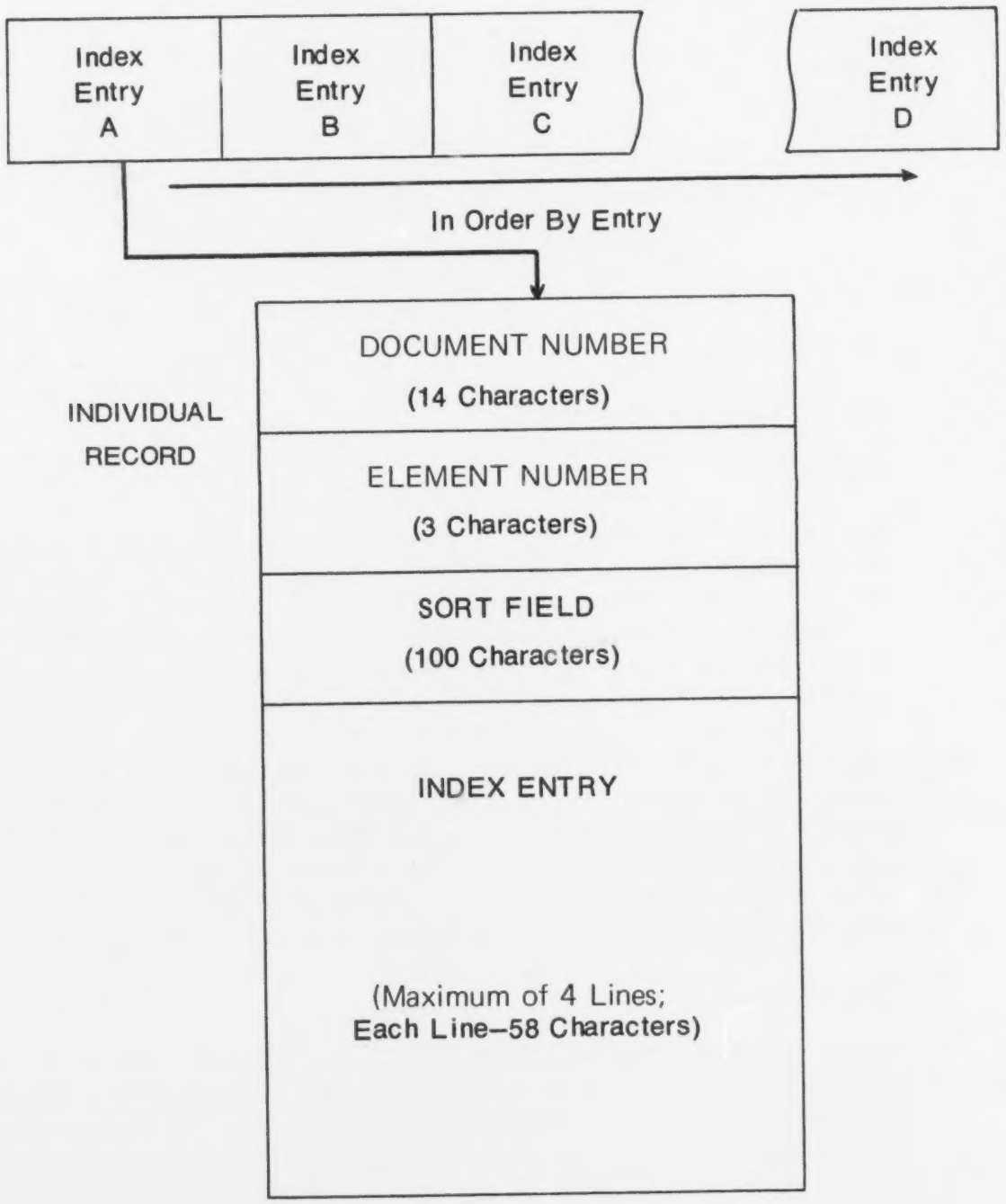
The subject descriptors from the content analysis portion of the reference are reformatted for inclusion in the Main Document Listing. The descriptive analysis data is reformatted by each bibliographic entry. As separate records in the Publication File, they may be aggregated on demand for the major listings and indexes to the listings. As a means of standardizing the form of each bibliographic entry, a sort of control field is built. This field will be the one used to determine the order of the various entries.

Programmer's Notes

Subject descriptors. The descriptors assigned to a document reference are of several types: format terms, geographic terms, date information and terms describing the subject contents of the document. Among the subject terms, there are two types: the broadly defined terms which are assigned as subject headings for the *Input Index* and the in-depth subject terms. It is the latter class of descriptor that is selected for inclusion on the Content Analysis File.

Sort control field. The sort control field of one hundred positions is a means of standardizing the bibliographic entries for sorting purposes. Quite often during data entry, the same information was presented with slight variations in spacing and punctuation which later caused problems in sorting. The sort control field eliminates this type of problem.

PUBLICATION FILE RECORD LAYOUT[1]



[1] Reproduced from Chapter VII of the *General Manual.*

Figure 8

A sort control field is developed only for Type '5' records (descriptive analysis data). In creating the publication record, the text portion of the input is moved into a work area. By means of a matrix, every position is tested individually. If the position contains a period, a comma, a hyphen followed by a blank, or a blank followed by a blank, it is not stored in the sort control field. If the position tested contains anything other than these special characters it is stored in the first available position of the sort control field. This process continues until either one hundred characters are stored or a new entry is reached. An example of how the text lines are stored in the sort control field follows:

Text Lines

LEXINGTON-FAYETTE CO., KY. PLANNING COM-
MISSION GEOGRAPHIC UNITS PLAN

Sort Control Field

LEXINGTON–FAYETTE CO KY PLANNING COMMISSION GEOGRAPHIC UNITS PLAN

P0015 – Sort the Content Analysis File

Abstract

The subject descriptors may be included in the Main Document Listing through the creation of a Content Analysis File. Since these descriptors may be numerous, it would be advantageous to print them in alphabetical order. The Content Analysis File is created in sequence by document number. This major arrangement will be retained while the sort additionally alphabetizes the descriptors within each document number reference.

P0025 – Sort the Publication File

Abstract

The Publication File is created from the CFS editing programs in sequence first by document number and then element number. For purposes of the Main Document Listing, the Publication File will be arranged primarily by document number, element number and entry. The sort control field will be used for sequencing the records by entry.

P0040 – Create the Subject Subset File

Abstract

The Major Subject Listing is organized so that each reference is cited once for each subject heading assigned to it. Included as part of each citation is the author and title information. P0040 scans the document reference to reproduce the selected information once for each citation. If a document contains more than the allowed maximum of ten subject headings, an error message is printed.

Programmer's Notes

Selection of the subset contents. Although the current contents of the Subject Subset File are author and title information, the choice of bibliographic elements may be changed without requiring additional programming. The lead card is the means by which the program selects the bibliographic data for inclusion in the file. The limit for the number of elements to be selected is fixed at ten.

Maximum number of subject headings per document reference. The average document reference has three subject heading assigned. The system has a limit of ten subject headings per reference. In the event that more than ten have been assigned, the additional entries will be listed for the document analyst's attention.

P0045 – Sort the Subject Subset File

Abstract

The Subject Subset File, created in document number and element number sequence, must be sorted to the sequence for the Major Subject Listing. At the conclusion of the sort, the file will be in sequence by subject heading, document number and element number.

P0050 – Print the Main Document Listing or the Major Subject Listing

Abstract

The Main Document Listing and the Major Subject Listing are printed by separate passes of this program. While the basic formats of the report are the same, the Main Document Listing may optionally include the subject terms used in the content analysis. The Major Subject Listing cannot include these terms. The Major Subject Listing is a variation and a subset of the primary citations in the Main Document Listing in which the entries are organized according to the assigned subject headings.

P0055 – Sort the Publication Index File

Abstract

To this point, the Publication File has previously been maintained in the same sequence as the Main Document Listing – namely, document number, element number and individual entry. For purposes of creating the indexes to the Main Document Listing, the sequence of the file must be by individual bibliographic entry. At the conclusion of the sort, the Publication Index File will be in sequence by individual bibliographic entry and document number.

P0060 — Create the Indexes to the Main Document Listing

Abstract

The creation of the indexes allows the user to produce one or several indexes to the Main Document Listing. The contents of each index section are selected and formatted according to the specifications stated in the input. The output media for the index sections may be printed copy, magnetic tape or both.

Programmer's Notes

Page length. The user provides P0060 with two page lengths: the printed page and the length of the tape image for double column indexes. These two page lengths are entered differently: the printed page through a carriage tape and the tape page length through a lead card. The two page lengths are usually, but need not be, the same.

Indexes on tape. The user may assign any of the indexes to be created on tape. However, the tape image sections have the restriction that the document numbers must appear in indented format under the index entry.

P0065 — Sort the Formated Page File of the Index Sections

Abstract

The creation of tape output for an index section is most prevalent for the double column indexes. When creating the tape, the entries are arranged in index, page, line and column number sequence. At the conclusion of the sort, entries will be in sequence by index, page, column and line (major to minor sequence). It is only by creating this type of tape output that both columns of a double-column page can be printed simultaneously.

P0070 — Print the Camera-Ready Copy of the Index Sections

Abstract

With the Formatted Page File by index sections, camera-ready copy of the double-column indexes can be produced. Since all formatting and selections of output options were determined in P0060, this is basically a tape-to-printer operation.

P0080 — Create Search-Publications Interface

Abstract

Special issues of the *Input Index* are most often cumulative issues of previous *Input Indexes* or specialized bibliographies. The latter are most easily located through a computer search. The search-publications interface was designed to compensate for the fact that CFS provided only printed output, namely the, *Retrieval Report*. The interface subsets and reformats the bibliographic information on the Selection File (the tape used

by the Search Module to print the *Retrieval Report*) or the Document Master File as a Special Publication File to be processed by the Publications Module.

Tape File Specifications

Publication File
Sorted Publication File
Publication Index File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:*
Fixed-length records of 409 characters
Blocking factor of 3 records
Padding record of 9s

Field	Size	Cols	Contents
Filler	X(58)	1- 58	
Document Number	X(14)	59- 72	
Element Number	X(3)	73- 75	
Sort Control Field	X(100)	76-175	
Number of Lines	X	176	
Text Lines	X(232)	177-408	Occurs as field of X(58). Maximum of 4 occurrences.
Record Mark	X	409	

Content Analysis File
Sorted Content Analysis File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:*
Fixed-length records of 38 characters
Blocking factor of 10 records
Padding record of 9s

Field	Size	Cols
Document Number	X(14)	1-14
Descriptor	X(23)	15-37
Record Mark	X	38

Subject Subset File
Sorted Subject Subset File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:*
Fixed-length records of 409 characters
Blocking factor of 3 records
Padding record of 9s

Field	Size	Cols	Contents
Subject	X(58)	1- 58	
Document Number	X(14)	59- 72	
Element Number	X(3)	73- 75	
Sort Control Field	X(100)	76-175	
Number of Lines	X	176	
Text Lines	X(232)	177-408	Occurs as fields of X(58).
Record Mark	X	409	Maximum of 4 occurrences.

Formatted Page File
Sorted Formatted Page File

a. *File Format:*
No header label
Data records, tape mark
No trailer label

b. *Data Record Format:*
Fixed-length records of 69 characters
Blocking factor of 10 records
Padding record of 9s

Field	Size	Cols
Index Number	X(3)	1-3
Page Number	X(3)	4-6
Line Number	X(3)	7-9
Column Number	X	10
Text Line	X(58)	11-68
Filler	X	69

References to the URBANDOC Final Report

Much of the information presented in this chapter is designed to be used with sections of the *General Manual* (G.M.) and other sections of the *Operations Manual* (O.M.).

For additional information on the Publication File, its format and the considerations in creating entries to the file, see:

Manual	Chapter	Section
G.M.	III – Document Analysis: Descriptive	Whole chapter
G.M.	VII – Systems Modules: Output	Publications Module, Publication File Description

For additional information on the design and goals of the Publications Module, see:

Manual	Chapter	Section
G.M.	VII – Systems Modules: Output	Publications Module, Function, Tasks

For additional information on operating this portion of the system, see:

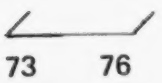
Manual	Chapter	Section
O.M.	VIII – Processing Cycles	Editing and Validation Cycle, *Input Index* File Maintenance Cycle, *Input Index* Production Cycle, Miscellaneous Publications Products
O.M.	IX – Operating Instructions	Publications Module
O.M.	XI – Tape Library and Report Controls	Publications Module
O.M.	XII – Timing	Publications Module

DATA ENTRY

Thesaurus Data Entry

Thesaurus Worksheet – Subject Descriptors, Subdescriptors
(See Figure 9.)

Accession Number.



Accession number is a four-digit code representing the term's sequence in the Thesaurus File. It is formed by the first letter of the term and a numerical sequence of the terms within the letter grouping. The numerical sequence is assigned with spaces of five or ten numbers in between depending upon the density of terms for each letter of the alphabet. When a term is deleted, its accession number becomes free for assignment to a new term.

URBANDOC has not yet had more than 999 terms for one letter of the alphabet. If this case should arise, some other scheme could be devised such as making the first two characters alphabetic.

This field is punched in every unit of each term set.

Term Section.

When the term is to be used as a descriptor, a pre-code must be entered in column 1:

'*': a precise descriptor (one which can be searched alone);

'#': a common descriptor (one which must be searched for in conjunction with a precise descriptor);

'-': an internal descriptor (one which can be searched alone but which is usually used for document numbers in descriptor form, etc.);

Descriptors will begin in column 2.

When the term is to be used as a subdescriptor, pre-code is not used and the term will begin in column 1.

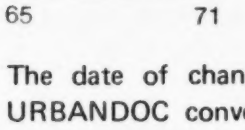
The formation of descriptors and subdescriptors and the valid characters allowed in the terms have been discussed in the *General Manual*. (See the *General Manual*, Chapter IV, specifically the Subject Thesaurus and also Chapter VI, specifically the Thesaurus File Description.)

Systems restrictions state that a term may be variable in length to fifty-three characters. However, when stored on the Document Master File, the terms used for content analysis are much shorter. (See Document Data Entry, specifically the Document Worksheet.) For this reason, URBANDOC has chosen to limit the length of its terms wherever possible.

Subdescriptors are usually between ten and fifteen characters in length. Descriptors are usually thirty-five to forty characters in length at most. Terms which may be assigned as subject headings in the *Input Index* have been limited to twenty-seven characters. The length of the subject heading was determined by the printed layout of the subject listing in the *Input Index*.

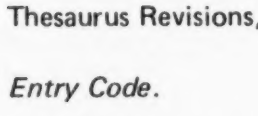
Each term is entered left-justified in the field, descriptors beginning in column 2 and subdescriptors in column 1.

Date of Entry.



The date the term is originally entered into the system takes the form of mm/dd/yy, where mm = month, dd = day and yy = year. The slashes must be included.

Date of Change.



The date of change reflects the most recent revision of a term's contents. As an URBANDOC convention, date of change is not used on this worksheet. It is used in Thesaurus Revisions, discussed later in this chapter.

Entry Code.

Entry code identified the type of transaction being processed:

'1': entry of a descriptor or subdescriptor;

'2': entry of a descriptor only (As an URBANDOC convention, this form is not used.);

'3': entry of a subdescriptor only.

Substitute Section. The CFS System made provision for preferred forms for a term used as a descriptor, subdescriptor or subject heading. The preferred form for the content analysis terms uses the search substitute. The subject heading uses the publications substitute.

The analyst must guarantee completeness of cross-referencing and check to see that the appropriate reverse entry has been made. When term B is listed for use in place of term A, there should be a main term for B in which there is an entry for A in the Used For Section.

Column 77 is precoded to keep the various entries within the term set in sequence.

Search Substitute

The search substitute is used for the preferred form for a term used as a descriptor or subdescriptor. A descriptor must be replaced by another descriptor and a subdescriptor by another subdescriptor.

When the main term is a descriptor, a pre-code must be entered in column 1 of the Search Substitute:

'*': a precise descriptor (one which can be searched alone);

'#': a common descriptor (one which must be searched for in conjunction with a precise descriptor);

The descriptor will begin in column 2.

The formation of descriptors as substitute terms is the same as for the main term. This has been discussed in the *General Manual*. (See the *General Manual*, Chapter IV, specifically the Subject Thesaurus and also Chapter VI, specifically the Thesaurus File Description.)

Each term being entered as a substitute must also appear in the Thesaurus as a main term. The term being entered here must agree with that form. Systems restrictions state that a term being entered as a substitute may be variable in length up to twenty-three characters. When the term used as a substitute is longer than twenty-three characters, the document analyst truncates the term after the twenty-third character so that the substitute and the main term will agree at least to twenty-three characters.

Although the CFS system allows for the substitution of subdescriptors, as an URBANDOC convention subdescriptors never have a substitute form.

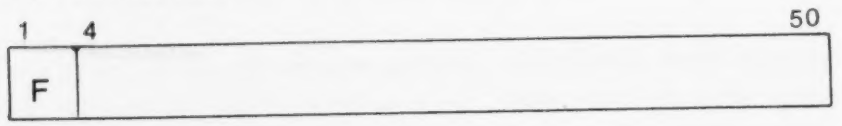
Subject Heading Substitute

The subject heading substitute is used to enter a preferred form for use as a subject heading in the *Input Index*. Although URBANDOC does not use the subject heading substitute, this section is described for other users.

The subject heading substitute would be used only for descriptors since subdescriptors would not be used as subject headings. Pre-code is not used with a subject heading substitute. The term begins in column 25 and may be variable in length to thirty-five characters.

Each term entered as a subject heading substitute must also appear as a main term in the Thesaurus. Although URBANDOC's subject headings are a maximum of twenty-seven characters, if the main term is longer than thirty-five characters the document analyst would truncate the term after the thirty-fifth character.

Used For Section.



When a term is used as the preferred form for another term, a record of this substitution must be kept in the main entry for the substitute. The entry specified in the Term Section has been used as the search substitute in the entries listed in the Used For Section.

The analyst must guarantee completeness of cross-referencing and check to see that the appropriate reverse entry has been made. When term B is listed as used in place of term A, there should be a main term for B in which there is an entry for A in the Used For Section.

Column 1 is precoded 'F', the notation for "used for". Each entry begins in column 4 and may continue to column 50. This field is large enough to contain any term entered as a main term in the Thesaurus.

Column 77 is precoded to keep the various entries for each term in sequence within the term set.

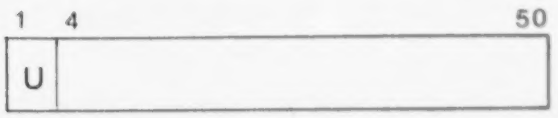
Cross-reference Section. The Cross-reference Section is composed of the various relationships that may exist between the main term being entered and other terms in the Thesaurus. Each term specified as a cross-reference must also be entered into the Thesaurus as a main term. The various relationships that may be specified are broader terms, narrower terms and related terms. (See the *General Manual*, Chapter IV, specifically the section on Subject Thesaurus.) This section also includes the Word Source Section and the Scope Note Section. The units of the Cross-reference Section may be identified by the code of '4' in column 80.

The total Cross-reference Section may not exceed sixteen lines of information. Be sure to use as much of the line as possible for the Word Source Section and the Scope Note Section. In all other sections there may be only one term per line.

In the event that a term must have more than sixteen units of Cross-reference Information, the term must be split and a continuation entry created. (See Chapter II, specifically the Thesaurus File.)

Column 77 is precoded to keep the various entries for each term in sequence within the term set.

Broader Term Section.

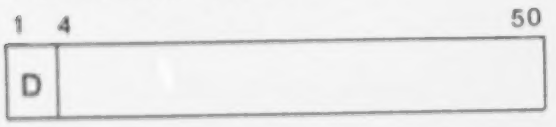


The term specified in this section is broader in meaning than the one used in the Term Section. The character in column 1, precoded 'U' on the worksheet, is the notation for a broader term. The entry begins in column 4 and may continue to column 50. This field is large enough to contain any term entered as a main term in the Thesaurus.

The analyst must guarantee completeness of cross-referencing and check to see that the appropriate reverse entry has been made. When term B is listed as a broader term for A,

there should be a main term B in which there is an entry for A as a narrower term.

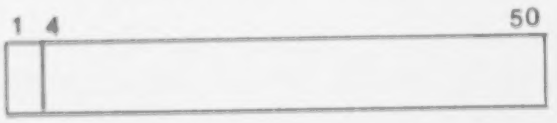
Narrower Term Section.



The term specified in this section is narrower in meaning than the one used in the Term Section. The character in column 1, precoded 'D' on the worksheet, is the notation for a narrower term. The entry begins in column 4 and may continue to column 50. This field is large enough to contain any term entered as a main term in the Thesaurus.

The analyst must guarantee completeness of cross-referencing and check to see that the appropriate reverse entry has been made. When term B is listed as a narrower term for A, there should be a main term for B in which there is an entry for A as a broader term.

Related Term Section.



The entries in this section display relationships to the descriptor or subdescriptor that are of a non-hierarchical nature. The term specified as the related term begins in column 4 and may continue to column 50. This field is large enough to contain any term entered as a main term in the Thesaurus. The document analyst must complete column 1 to indicate the various types of relationships, which are as follows:

C: term following is to be considered for use in co-ordination with the descriptor;

M: term following is a subdescriptor to be considered for use with the descriptor;

O: term following is opposite in meaning to the descriptor;

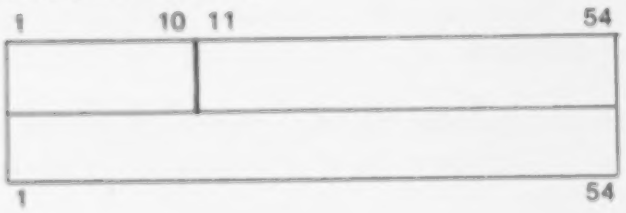
P: term following is prohibited from use in conjunction with the descriptor;

R: occurs only in cases where the prime term is a subdescriptor and all S terms are suggested descriptors which it may modify.

The 'O' and the 'P' relationships are used less frequently than any of the others.

Reverse relationships must also be maintained for these entries.

Word Source Section.

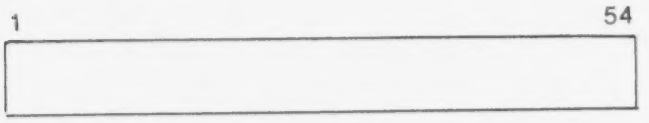


The date on which the term was discovered is entered into columns 1-10. The date of the source takes the form mm/dd/yy, where mm = month, dd = day and yy = year, when all three elements are available. The two remaining columns (9-10) are used to separate date from the word source when the file is printed.

The source of the term is entered in columns 11-54 of the first line. If necessary, columns 1-54 of the next line are also available. This field could contain the name of another organization, file, individual, thesaurus, etc.

Column 77 is precoded to keep the various entries for each term in sequence within the term set.

Scope Note Section.



This section is used only when it is necessary to express an adequate definition of the term, indicating its limitations of usage or indicating that this term is no longer in use as of a particular date. This last note is a guide for search purposes facilitating retrospective searching of the file.

i.e. CIP
F CAPITAL IMPROVEMENT PROGRAMS
S SEARCH ON CIP FOR DOCUMENTS
INDEXED UNDER THIS TERM PRIOR TO
MARCH 1969

Column 77 is precoded to keep the various entries for each term in sequence within the term set.

Terms and Symbols used in the Hard-Copy Thesaurus

D	Term Usage	Appears in the left margin to indicate entry is a descriptor.
S	Term Usage	Appears in the left margin to indicate entry is a subdescriptor.
B	Term Usage	Appears in the left margin to indicate entry may be used as either a descriptor or a subdescriptor.

At the present, only 'B' and 'S' are being used. All 'B' entries are being used only as descriptors. The programming system allows for a 'B' to be used in either way, but the project has limited itself to only a descriptor usage.

*	Pre-code	Precedes all descriptors indicating its search status. A subdescriptor is not preceded by any code.

USE	Substitute Section	The word 'USE' indicates that the term following is to be used in place of the entry. This is a *see* reference. *ACCIDENTS — Do not use ACCIDENTS. USE *SAFETY — Use instead the term SAFETY.
F	Used For Section	The word following 'F' is a synonym, which has been replaced for searching by the preferred term.
U	Up-post Term	Term, broader in meaning, to which the entry could be up-posted.
D	Down-post Term	Term, narrower in meaning, to which the entry could be down-posted.
C	Co-ordinate Term	Term which is to be considered for co-ordination between the named descriptor and the entry.
M	Modifying Term	Subdescriptor which is to be considered for modification of the entry. *COMPENSATORY EDUCATION M–AFTER–SCHOOL
O	Opposite Term	Term which has the opposite meaning of the entry, used infrequently.
P	Prohibited Entry	Use of the entry is prohibited with the named descriptor or subdescriptor, used infrequently.
R	Related Term	This is a descriptor related to the entry.
S	Modifying Term	This is a descriptor listed under a subdescriptor entry to show modification. It shows in reverse the use of the symbol 'M'. AFTER–SCHOOL S – COMPENSATORY EDUCATION

In making use of the URBANDOC Thesaurus, the reader should remember that the Thesaurus, in its present form, serves to assist in the processing of documents for input into the computer. The Thesaurus functions as an authority and a guide to the document analyst and provides the basis for machine query. Under each entry is gathered all relevant terms which reflect the scope and relationships of the entry.

THESAURUS WORKSHEET – SUBJECT DESCRIPTORS, SUBDESCRIPTORS

ACC. NO. _____
73 76

TERM SECTION

1 2	term	54	55 date of entry 62	65 date of change 71	80 entry code
			/ /	/ /	

SUBSTITUTE SECTION

1 2	24

SUBJECT HEADING SUBSTITUTE

25	60	77

USED FOR SECTION

1 4	50	77	80
F		A	4
F		B	4
F		C	4
F		D	4
F		E	4

BROADER TERMS

1 4	50	77	80
U		F	4
U		G	4
U		H	4

NARROW TERMS

1 4	50	77	80
D		I	4

Figure 9

THESAURUS WORKSHEET – SUBJECT DESCRIPTORS, SUBDESCRIPTORS

(Continued from preceding page)

RELATED TERMS

C – Consider co-ordination with the named descriptor.

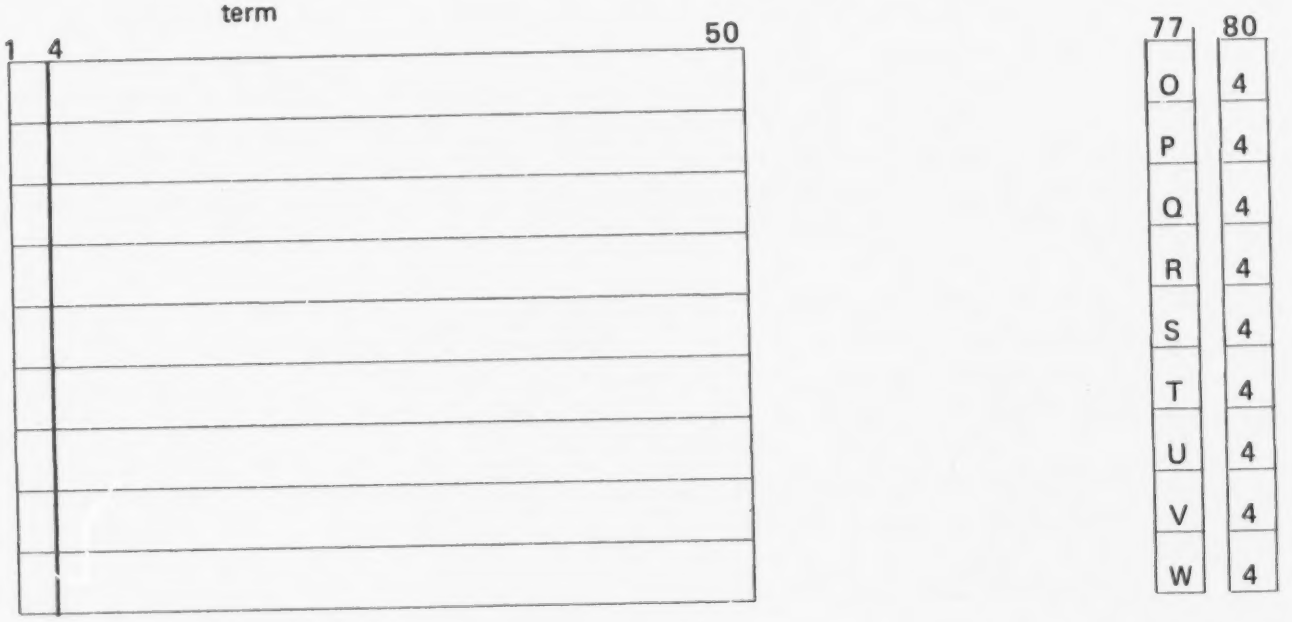
M – Consider modification by the named subdescriptor.

O – Entry is opposite in meaning to the descriptor or subdescriptor named in the term entry.

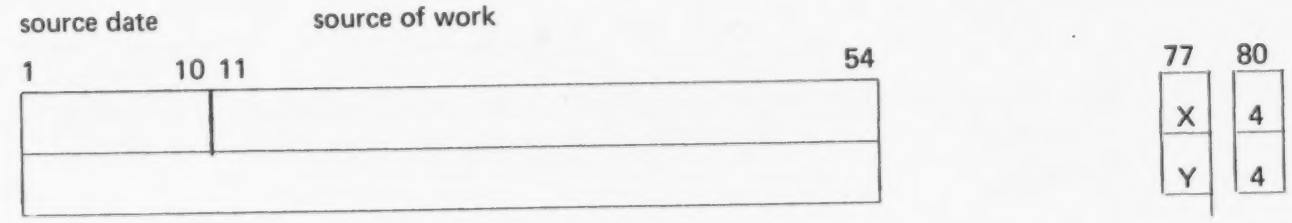
P – Entry is prohibited from use with the names descriptor or subdescriptor.

R – Related term.

S – Consider the entry as a modifier to the names descriptor.



WORD SOURCE



SCOPE NOTES

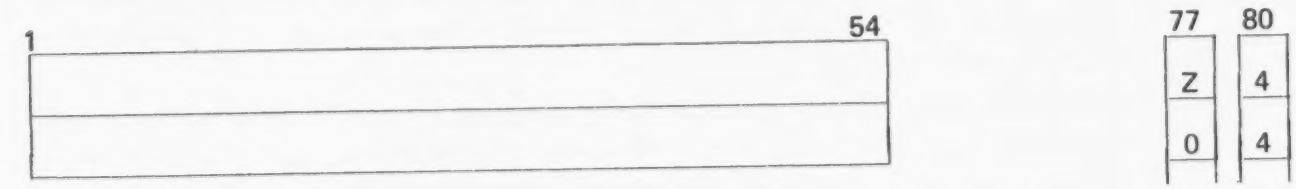


Figure 9 (cont.)

THESAURUS WORKSHEET – GEOGRAPHIC DESCRIPTORS

(See Figure 10.)

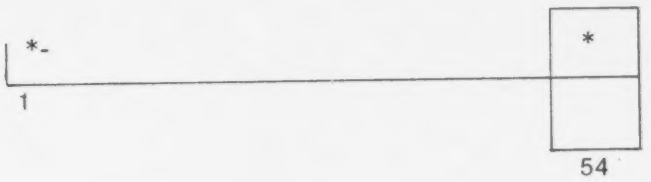
The geographic worksheet is based on the techniques developed by URBANDOC for handling the geographic portion of the document analysis. As mentioned in the *General Manual*, numerical codes rather than natural language terms were chosen for the geographic analysis. The geographic descriptors are based on the Census Bureau Code which enabled URBANDOC to combine state, country and place into one descriptor. Other codes also exist for Standard Metropolitan Statistical Area (SMSA) and city size (by ranges of population).

The CFS system made provision for the substitution of a preferred term for the natural language form used in the input. In tailoring the system to URBANDOC's use, a choice had to be made between entering the code or the natural language term as an indexing term (in which case the system would make the replacement). The latter method was selected for the sake of error detection and correction. It is easier to detect a keying error in a word than in a numerical code. In addition, computer validation of descriptors would not detect a transposition error that yielded a valid code.

This form was designed as a collective worksheet for all the different entries related to a single place name. The worksheet is divided into three sections: place name to code forms, Census Code for place name and SMSA code for place name. Broader, narrower and related terms, as well as word source notes, do not occur in this worksheet.

Place Name Section. This section contains all the data for entering the place names and their corresponding code forms.

Place Name Term.



Since the geographic terms are all descriptors, a pre-code must be entered in column 1. Geographic terms (except for #GEO) are all precise descriptors and as such use the '*' pre-code.

The geographic descriptor, when a place name term, is basically an alphabetic term. When the Thesaurus is maintained on tape, all the terms are merged together in term sequence. In order to separate the geographic place names from the subject descriptors and subdescriptors, a block code is entered in column 2. The block code '-' serves to group all the place names together.

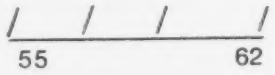
The formation of place name descriptors is the same as for the subject descriptors and subdescriptors. These have been discussed in the *General Manual*. (See the *General Manual*, Chapter IV, specifically the Subject Thesaurus and Geographical Thesaurus and also Chapter VI, specifically the Thesaurus File Description.)

The place name begins in column 3 and may continue through column 54. Systems restrictions allow a term to be variable to fifty-three characters. Since the block code counts as one character, the actual place name itself cannot exceed fifty-two characters. During the content analysis, the analyst will use the natural language form.

When the '*' and/or '/' is checked by the analyst, it is an indication that the term is to be entered once for each checked form. In these variations, the '*' or '/' takes the place of the '.' terminal punctuation. The '*' indicates an SMSA entry for the place name, the '/' a city size entry. Each variation of the place name is a unique term on the Thesaurus and will receive a distinct accession number.

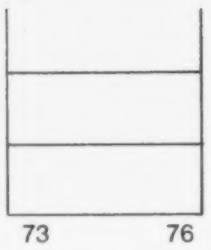
In cases of foreign place names and non-Census coded U.S. place names (rivers, regions, new towns, etc.), only the first line of the worksheet is used. In cases of regional types of entries, the scope note section is also used.

Place Name Date of Entry.



The date of entry is the date of the term's original entry into the system. It takes the form of mm/dd/yy, where mm = month, dd = day and yy = year. This field will apply for each variation of the place name.

Place Name Accession Numbers.



Each geographic descriptor is assigned a unique accession number, differing in format and formation from the subject term accession numbers. While there are further differences which separate the SMSA descriptors' accession number from the other geographic descriptors, these differences will be discussed separately.

The accession number is four digits, constructed to allow for a large number of entries. It is to be expected that the geographic terms will far outnumber the subject entries. The first two characters are alphabetic, beginning with AA and running through ZZ. Starting with AA00, each entry would increase by one: AA01-AA99, AB00-AB99, etc. Each entry on the worksheet receives the next available number. This coding scheme applies to all geographic descriptors, both domestic and foreign, except for SMSA.

A geographic descriptor contains basically coded data. As such it does not expand and there is never any need to refer back to their accession numbers. Should a code change and it is necessary to delete the term, it is not necessary to enter the original accession number. The original input data is not needed.

The main term is recorded and assigned the next available accession number. To be sure of the next available number, the analyst should note the last one assigned.

This field is punched in every unit of a term set.

Place Name Entry Code.

/ 1 /
80

Entry code identifies the type of transaction being processed. Since place name is always a descriptor and it is being added to the Thesaurus, this field has been precoded '1' on the worksheet.

Place Name Search Substitute. Three search substitutes are provided. Only one form may be used in a term set but each place name variation uses a different form. Each substitute receives the corresponding accession number.

Since place name terms are descriptors, the substitutes must also be descriptors. As such, they require a pre-code in column 1. All geographic terms are precise descriptors and use the '*' pre-code.

Each term being entered as a substitute must also appear in the Thesaurus as a main term. The term being entered here must agree with that form. If the term is not already on the Thesaurus, it may be entered under the Census Code Section or SMSA Section of the worksheet.

Column 77 has been precoded to keep the various entries for each term in sequence within the term set.

Census Code Substitute.

/ *
1 24

The Census Code substitute corresponds to the place name form ending in '.'. Since the Census Code is numeric, the code will not be merged in with the subject descriptors and subdescriptors and no block code is required. The Census Code is entered beginning in column 2.

SMSA Substitute.

/ * ØSMSA- /
1 24

The SMSA substitute corresponds to the place name form ending in '*'. Since SMSA is partially alphabetic, the codes would be merged with the subject descriptors and subdescriptors. Therefore a block code is required. Both the SMSA block code of '0' and the 'SMSA' root of the term have been precoded on worksheet. The user need only supply the remainder of the term.

By referring wherever feasible to SMSA, it is possible to reduce the number of place name references that need be made. Cross-referencing is not made from the SMSA to the region. It is made from the county or town if that county or town is part of the SMSA.

City Size Substitute.

/ *CITY SIZE- /
1 24

The city size substitute corresponds to the place name form ending in the '/'. Although city size is partially alphabetic, there are only sixteen codes for city size and it was not judged necessary to segregate these from the subject descriptors and subdescriptors. The city size root of the term has been precoded on the worksheet. The user need only supply the remainder of the term.

Place Name Scope Notes.

1 54

There is provision for a scope note (entered in columns 1-54) with a maximum of sixteen lines. These scope notes are restricted to use with regional types of entries where URBANDOC, through the use of source documents, provides a definition of a region.

Types of Scope Notes

Appalachia
Includes part of Alabama, Georgia, Kentucky, Maryland, New York North Carolina, Ohio, Pennsylvania, South Carolina, Virginia and West Virginia.

Appleton, Wis.
Use *350550080
Included in Fox Valley Wisconsin Region.

Southeast Michigan Metropolitan Region
Includes 0SMA-0450, 0SMA-0080, Monroe Co., Saint Clair Co., and Livingston Co., as defined by Southeast Michigan Council of Governments in ODERTRSM68RC,

Column 77 has been precoded to keep the units of a term set in sequence. The code for cross-reference data has already been entered on the worksheet for column 80. In the event that more than sixteen cross-references should be entered for a term, the term set must be split and a continuation record created.

Census Code Section. This section contains the correspondence of the Census Codes to the individual place names.

Census Code Term.

/ * /
1 24

The Census Code term will correspond exactly to the term entered as the Census Code substitute.

Census Code Date of Entry.

/ / / /
55 62

The date of entry will correspond exactly to the date of entry entered under the Place Name Section.

Census Code Accession Number.

/ /
73 76

The accession number is assigned the same way as under the Place Name Section. Assign the next available number.

Census Code Entry Code.

/ 1 /
80

Entry code corresponds to the entry code under the Place Name Section.

Census Code Used For Section.

F/-
1 4 54

This section provides for the reverse entry for place name. It will correspond to the place name form of the term ending in '.'. The 'F' precoded on the worksheet is the notation of a "used for" term. The '-' block code has already been precoded onto the worksheet in column 4. The place name should be entered starting in column 5.

Column 77 has already been precoded to keep the units for a term set in sequence. The code for the cross-reference data has also been precoded for column 80.

SMSA Section. This section contains the correspondence of SMSA codes to the individual place names.

SMSA Term.

/ *ØSMSA- /
1 24

The SMSA provides the reverse entry for the SMSA substitution. Before creating the SMSA entry, the analyst must check to see if the SMSA term already exists. If it does, the new entry causes the SMSA to be treated as a changed term and as such follows the rules for the subject term. The SMSA series is treated differently from the other geographic terms since it is frequently expanded to include additional place name entries. The other terms are usually entered and not changed.

If the SMSA term does not exist, the analyst will complete this section of the worksheet. This form will correspond exactly to the SMSA substitute entered above.

SMSA Date of Entry.

/ / / /
55 62

The date of entry will correspond exactly to the date of entry entered under the Place Name Section.

SMSA Accession Number.

/ Ø /
76

The accession numbers for SMSA are the actual SMSA number codes. This is possible since they are both four digits. There is no danger of duplicating the accession number since this is the only term with an entirely numeric accession number. The SMSA series is treated differently since it is frequently expanded to include additional place names. For this reason, it is necessary to know the original accession number entry.

This field is punched in every unit of each term set.

SMSA Entry Code.

/ 1 /
80

Entry code corresponds to the entry code under the Place Name Section.

SMSA Used For Section.

```
/F/-                                                    */
1  4                                                    54
```

This section provides for the reverse entry for place name. It will correspond to the place name form of the term ending in '*'. The 'F' precoded onto the worksheet is the notation of a "used for" term. The '-' block code has already been entered in column 4. The place name should be entered starting in column 5. The terminal punctuation character '*' has been precoded onto the worksheet.

Column 77 is precoded to keep the units for a term set in sequence. The code for the cross-reference data has also been precoded for column 80.

By referring wherever feasible to SMSA, it is possible to reduce the number of place name references that need be made. Cross-referencing is not made from SMSA to the region. It is made from the county or town if that county or town is part of the SMSA.

City Size Cross-reference. As an URBANDOC convention and economy measure, city size reverse substitutions are not entered.

Thesaurus Revisions

During the project, it became apparent that it was not necessary to retain the original Thesaurus Worksheet. A Change Worksheet for subject terms was developed to cover any requisite changes to existing terms of the Thesaurus File. This sheet can be discarded after the file has been updated correctly. The nature of the changes to the existing terms are generally the addition and/or deletion of related term references or the deletion of an entire term. New terms are treated by the original entry procedures.

Basically, the Change Worksheet is an abbreviated version of the Thesaurus Worksheet, with the elimination of all coding. The keypunch operator uses a blank Thesaurus Worksheet as her model for the coding. Before punching the changes, the term set is pulled from the file by using the Thesaurus Shelf List (a report of each term and its corresponding accession number).

For each revised term, a delete card must be prepared. This card will remove the existing form of the term from the Thesaurus File. The delete card is identical to the term card except that it includes change data and a different entry code in column 80.

For a term with an entry code of '1', the delete card must be coded 'J' in column 80.

For a term with an entry code of '3', the delete card must be coded 'L' in column 80.

The change data receives the date in the form mm/dd/yy.

THESAURUS WORKSHEET – GEOGRAPHIC DESCRIPTORS

PLACE NAME SECTION.

PLACE NAME

*-
1

*
/
54

ENTRY DATE

/ /
55 62

ACCESS NOS.

73 76

1
80

CENSUS CODE

*
1 24

/
77

*@SMSA-
1 24

/
77

*CITY SIZE-
1 24

·
77

SCOPE NOTES

1 54

0 4
1 4
77 80

CENSUS CODE SECTION.

CENSUS CODE

*
1 24

/ /
55 62

73 76

1
80

F /-
1 6 24

A 4
77 80

SMSA SECTION.

*@SMSA-
1 24

/ /
55 62

@
73 76

1
80

F /-
1 6

*
54

A 4
77 80

Figure 10

In the case of a deleted term, the creation of a delete card is all that is necessary in the way of machine-readable input. The term set may be discarded.

For a term being revised, whether information is being added and/or deleted, the term set must be modified according to the Change Worksheet. Only changes to a term are noted on these new worksheets. The deletion of information is accomplished by removing the cards from the set. The units to be added are punched with the correct sequence number in column 77. The sequence number for new units is selected from the next available number in the set. After the changes have been made, the term set is reassembled.

When preparing the input for the actual update of the Thesaurus File, all the delete cards are grouped together. The new and revised term sets follow. A pre-list of the Thesaurus Input is prepared for code validation and proof-reading. After the Thesaurus File has been updated, the delete cards may be discarded and the revised term sets returned to the file.

Program Control Card – Thesaurus Worksheets

Column	*Contents*
1	'1'
2-3	'A's
4	'1'
5-54	'A's
55	' '
56	'&'
57	'1'
58-62	'A's
63	' '
64-72	'&'s
73	'1'
74	' '
75-76	'&'s
77	'1'
78	' '
79	'&'
80	' '

Document Data Entry

Document Worksheet
(See Figure 11)

The Document Worksheet follows the pattern of the punched card in that each line on the worksheet represents one unit to be punched. However, such information as responsibility for work done, date keypunched, etc. is not entered as part of the machine-readable record. The worksheets are retained and filed for backup until the reference has been entered onto the Document Master File.

The Document Worksheet, like the Thesaurus Worksheets, have been preceded wherever

possible to minimize the professional time that must be spent in completing the data entry for a reference. This format has been modified from the basic CFS format to minimize the information that must be manually entered. Wherever possible, computer generation of standardized entries and computer insertation of CFS coding is performed.

Responsibility for Work Section.

descriptive analysis _____

content analysis _____ initials of appropriate staff member

keypunching date _____

This information is not entered as part of the machine-readable input for a reference.

Routing Section

publications II # _____ designation of which *Input Index* document is assigned

retrieval _____ if document is for retrieval only

(See *Demonstration Report*, Chapter IV, specifically the Currency Statement and Chapter V, specifically Mix of the Document Base.)

This information is not entered as part of the machine-readable input for a reference.

Document Number.

This field is punched for every unit in the reference. (See the *General Manual*, Chapter II, specifically Document Identification.)

Entry Code.

This section has been provided for the addition of a new reference to the Document Master File. This field is punched for every unit in the reference.

Format Term Section.

15 18	20	23 46
0060	2	*UPAP REPORT
0070	2	#ARTICLES
0080	2	#SPECIAL REPORTS
0090	2	*PERIODIC REPORTS
0100	2	*
0110	2	*

Unit numbers (columns 15-18) 0060-0110 are assigned to format terms. Usually only one format term is assigned per document. The most common types of formats, units 0060-0090, have been precoded on the worksheet. For these terms, the analyst need only check the correct term in column 20. Units 0100-0110 are left open for format terms other than those already entered on the worksheet.

Date Analysis Section.

15 18	20	23 49			
0130	2	#DATES 19--	19----	1---	1---
		pub-lish	entry	content single or upper	content lower

Unit 0130 is dedicated to date analysis information. Three types of dates may be entered here: the date of publication of the document; the date the document was processed by the document analyst; the time period under discussion in the document, either retrospective or projected.

The date of publication takes the form yyyy, where yyyy = full year. The worksheet has been precoded for the first two digits of the year.

The entry date takes the form yyyymm, where yyyy = full year and mm = month. The entry date has been precoded for the first two digits of the year.

The content date(s) takes the form yyyy, where yyyy = full year. The first digit of the year has been precoded. For a date later than 1999, the analyst can override the precoding. When the content date takes the form of a range, the upper range is entered in the first content date field. The lower date of the range is entered in the second content date field.

Geographic Analysis Section.

15 18	20	21 22	23 58
0170	2		#GEO
0180	2	/*	*-
0190	2	/*	*-
0200	2	/*	*-

More precoding is done in these units because of the complex nature of the geographic entries. The programming system has both an expansion capability (Pre-edit Module) and a substitution capability (File Maintenance Module). The document analyst enters the natural language place name term only once, checking the '/' and '*' in columns 21-22 where appropriate. Programs exist which use these codes to create the separate city size and SMSA entries. (See Chapter III, specifically program E0020.) The programming system will later substitute numeric codes for the natural language terms. (See Chapter IV, specifically programs MAINT2 and MAINT3.)

The '*-' in columns 23-24 are precoded on the worksheet and are part of the geographic terms. As discussed under Thesaurus Worksheet – Geographic Descriptors, these characters separate the geographic place name terms from the subject terms on the Thesaurus File. The remainder of the place name is entered starting in column 25. (See Thesaurus Worksheet – Geographic Descriptors. See also the *General Manual*, Chapter IV, Document Analysis: Content, specifically General Considerations and the Geographical Thesaurus.)

If a term used in indexing does not appear in the Geographic Thesaurus, a Geographic Worksheet must be completed for this entry. (See Thesaurus Worksheet – Geographic Descriptors.)

Though the sample shows unit numbers 0170-0200, the Geographic Analysis Section may range from units 0170-0490.

Content Analysis Section.

15 18	20 22	23 35	58	59 70
0510	2	*		
0520		*		
0530	2	*		
0540		*		

The content analysis section ranges from units number 0510-0790. It is here that the descriptors and subdescriptors describing the contents of the document are entered. Unlike the Geographic Analysis Section, if a term contains a preferred form, the document analyst enters the preferred form onto the worksheet. (See the *General Manual*, Chapter IV, Document Analysis: Content, specifically General Considerations.)

Each descriptor must be entered in a line precoded '2' in column 20. This is the code for descriptors (as opposed to subdescriptors). It is entered beginning in column 24. Even though the term may be quite long on the Thesaurus File, when entering it onto the Document Worksheet, it cannot be longer than thirty-five characters. This presents no problem for the current URBANDOC Thesaurus File since all descriptors already in the file end at the thirty-fifth character. If new terms are added which are longer than thirty-five characters, when entering them onto the Document Worksheet as an indexing term they must be truncated at the thirty-fifth character (column 58). It is recommended that new terms subscribe to the length rule.

If a particular descriptor is selected as a subject heading entry for the *Input Index*, it may so designated by entering 'SH' in columns 21-22 of the unit in which the descriptor is entered. These descriptors will be used to create the subject listing of the *Input Index* and will also appear in the citation of the reference in the Main Document Listing.

Subdescriptors may also be assigned to a descriptor. (See Chapter IV, specifically the Document Master File.) On the Document Worksheet, a subdescriptor may not be longer than twelve characters. Subdescriptors are entered in full on the Thesaurus File. However, when used as a term in content analysis, it is truncated after the twelfth character, i.e., STATE–MUNICIPAL is entered as STATE–MUNICI.

The first subdescriptor is entered in the same unit as the descriptor. It will appear in columns 59-70. The second and other descriptors will be entered in following units. For these subdescriptors, the entry will be made in columns 35-46 of the unit. Only one subdescriptor may be entered per unit.

Each of the additional subdescriptor units must be coded '3XX' in columns 20-22, the '3' being the code for a subdescriptor unit and the 'XX' indicating that the descriptor being modified is in the input.

The Content Analysis Section of the worksheet has alternate lines precoded for descriptors ('2' units). This allows the document analyst to assign two subdescriptors per descriptor. If more than two subdescriptors must be assigned, the document analyst may override the precoding in the following unit.

Any modifiers assigned to a descriptor coded 'SH' (*Input Index* subject heading) are ignored for subject heading purposes.

Free Text Section.

15 18	20 22	23 26 80
2030	5	
2040	511	
2050	511	

3010	521	
3020	5	

Units 1010-5040 are devoted to free text segments (descriptive analysis information). Some of the bibliographic entries are standard for every document. These entries have element number precoded for the document analyst. For the variable elements, element number must be completed by the document analyst.

The CFS system made no provision for formatting and editing bibliographic information and all information is stored as entered in the worksheet. URBANDOC has established the following conventions for entering data. Except for element '23', all bibliographic entries start in column 23 of the unit. If it is necessary to continue an entry to a succeeding unit, the continuation line begins in column 26. Since element '23', the abstract and notation of content, usually takes paragraph form the conventions are reversed. The first line of a paragraph begins in column 26. All continuation lines for that paragraph begin in column 23. (See the *General Manual*, Chapter III, Document Analysis: Descriptive.)

The document analyst must be sure to enter all bibliographic entries in numerical sequence by element number. Although URBANDOC has not found it necessary to do so, the document analyst can use the intervening unit numbers for additional information (i.e., 2041-2049). The Descriptive Analysis Section may also be extended to unit 9999.

Editing and Validation

Editing and Validation includes the computer detection of some errors and the production of listings to visually detect others. The Pre-edit Module performs the bulk of the computer error detection as well as expanding the geographic descriptors, date analysis, imprint for subdocuments and subject heading entries. The sequence of the input, both by document number and unit number, is checked. Each reference is examined for the presence of main author, title, imprint and collation entries and at least one subject heading entry. Format and coding within the input are also checked for validity and consistency. The File Maintenance Module verifies the descriptors, subdescriptors and subject headings for spelling, term length, authorized usage, etc.

Various error schedules are produced with codes for each error to allow the document analyst to correct the units. The various error listings indicate the document number and

DOCUMENT WORKSHEET

document number

1 14

entry code: A (19)

descriptive analysis _____

content analysis _____

keypunching date _____

Routing:

publications II # _____

retrieval _____

59 70

15 18	20	23 46
0060	2	*UPAP REPORT
0070	2	#ARTICLES
0080	2	#SPECIAL REPORTS
0090	2	*PERIODIC REPORTS
0100	2	*
0110	2	*

			pub-lish	entry	content single or upper	content lower
0130	2	#DATES	19--	19----	1---	1---

15 18	20	21 22	23 58
0170	2		#GEO
0180	2	/*	*-
0190	2	/*	*-
0200	2	/*	*-
0210	2	/*	*-
0220	2	/*	*-
0230	2	/*	*-
0240	2	/*	*-
0250	2	/*	*-
0260	2	/*	*-
0270	2	/*	*-

Figure 11

DOCUMENT WORKSHEET

(Continued from preceding page)

15 18	20 22	23 35 58	59 70
0510	2	*	
0520		*	
0530	2	*	
0540		*	
0550	2	*	
0560		*	
0570	2	*	
0580		*	
0590	2	*	
0600		*	
0610	2	*	
0620		*	
0630	2	*	
0640		*	
0650	2	*	
0660		*	
0670	2	*	
0680		*	
0690	2	*	
0700		*	
0710	2	*	
0720		*	
0730	2	*	
0740		*	
0750	2	*	
0760		*	
0770	2	*	
0780		*	
0790	2	*	

Figure 11 (Cont.)

DOCUMENT WORKSHEET

(Continued from preceding page)

01 personal author
02 corporate author
03 anonymous
— MAIN ENTRY

04 acronyms
05 joint personal author
06 joint corporate author
07 consultant
08 misc. corporate name
09 misc. local place name
— ADDED ENTRY

10 corporate author name x ref.
11 title
13 distinctive series title
15 meaningless title
21 imprint and collation
23 abstract & notation of content
25 UPAP project number
31 legal citations
36 literature citations
56 acquisition information
97 geographic index name

15 18	20 22	23 80
1010	5	
1020	5	
1030	5	
1040	5	
1050	5	
2010	5	
2020	5	
2030	5	
2040	511	
2050	511	
3010	521	
3020	5	
3030	5	
3040	5	
3050	5	
4010	525	
4020	5	
4030	5	
4040	5	
5010	5	
5020	5	
5030	5	
5040	597	

Figure 11 (cont.)

unit number of the entries in error. The Formatted Listing provides hard-copy for the input after processing by the Pre-edit Module. It is then possible for the analyst to use this listing as the master copy for all the necessary corrections.

Following the production of the Formatted Listing, a pre-publication sample of the *Input Index* is produced. This is used concurrently with the Formatted Listing as an additional editing tool since, except for subject headings, the bibliographic information is not edited or validated by the computer system. This version of the *Input Index* appears in the same format and arrangement as the final copy. These procedures allow the document analyst to visually check for spelling, form and consistency among the bibliographic entries. These are particularly apparent when entries are seen grouped together with others of the same type.

The *Input Index* consists of the Major Subject Listing, the Main Document Listing, listings of 701 reports, corporate and project names, consultants, personal names, geographic listing, significant titles and statutory citations. In addition, special segments (imprint and collation and accession information) are listed to verify consistency in entry form.

The specialized listings should be proof-read with the following guidelines in mind:

Major Subject Listing. The Major Subject Listing provides some measure of the consistency in the use of the headings displayed in the document array. It can also serve as a check on the relatedness of the heading to the title since the title appears with the heading in the listing. It may also seem desirable at this point to add other more generic terms to some documents in order to group similar material under a single heading.

Corporate and Project Name Index. The editing programs do not reconcile the differences in spelling and word order of such ihformation. Even though the corporate authority list is used as a reference in assignment of the entry, errors do occur and the same entry when entered differently will appear scattered in the listing. i.e., LABOR DEPT. and LABOR DEPARTMENT OF NEW YORK CITY. COMMUNITY RENEWAL PROGRAM and NEW YORK. CITY. CITY PLANNING COMMISSION. COMMUNITY RENEWAL DEPT.

Personal Name Index. URBNADOC does not maintain strict control over the forms for personal names. They are entered as found on the document in keeping with the current practice in the information systems community. It is possible, however, to pull together scattered entries for a personal name, if deemed necessary.

Imprint Listing. This printout provides an opportunity to check on inconsistencies in form and content of the imprint.

Significant Title Listing. Spelling errors in titles are most readily apparent here.

Geographic Index. Spelling and consistency in this listing are checked.

Statutory Citations. Spelling and validity of entry form are checked.

All detected errors are noted and transferred to the Formatted Listing. The keypunch

operator then has a single master copy from which to punch all the necessary corrections.

Besides corrections of detected errors, it sometimes becomes evident that additional units of information may be necessary to complete a document record. These additions can be added at this time to the Formatted Listing, inserting the new information with all its correct coding.

Example 1: Adding information

Formatted Listing before addition

Document #	*Unit #*	*Code*	*Type*	*Text*
BCOUNSG68WP	0530	A	2	*WATER RESOURCES
BCOUNSG68WP	0550	A	2	*WATER POLLUTION

Form of input (for addition)

BCOUNSG68WP	0540	A	2	*WATERFRONTS

Formatted Listing after addition

Document #	*Unit #*	*Code*	*Type*	*Text*
BCOUNSG68WP	0530	A	2	*WATER RESOURCES
BCOUNSG68WP	0540	A	2	*WATERFRONTS (unit added)
BCOUNSG68WP	0550	A	2	*WATER POLLUTION

A revised entry automatically replaces the original entry and inserts the desired information. All that is needed to a change to a unit is an input card with the same unit number.

Example 2: Changing information

Formatted Listing before change

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
DDREGCL69LB	1040	A	5	05	ETTER, ORVAL
DDREGCL69LB	2040	A	5	11	LOCAL GDUNDARIES
DDREGCL69LB	3010	A	5	21	1969, 49P.

Form of input (for change)

DDREGCL69LB	2040	A	5	11	LOCAL BOUNDARIES

Formatted Listing after change

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
DDREGCL69LB	1040	A	5	05	ETTER, ORVAL
DDREGCL69LB	2040	A	5	11	LOCAL BOUNDARIES (unit changed)
DDREGCL69LB	3010	A	5	21	1969, 49P.

Data that is to be completely eliminated is coded with a 'D' in column 19. If just an individual unit is to be deleted, the input card contains the document number, unit number, 'D' in column 19 and type code in column 20. If an entire document is to be deleted, the input card contains the document number, 'D' in column 19, and type code in column 20. Only one input card is required to delete an entire document from the file.

Example 3: Deleting a unit

Formatted Listing before deletion of a unit

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
GGEORPB69RS	9010	A	5	99	6APPBS
GGEORPB69RS	9020	A	5	99	6BPROGRAMMING
GGEORPB69RS	9030	A	5	99	6CADMINISTRATIVE ORGANIZATION

Form of input (for deletion)

GGEORPB69RS	9020	D	5		

Formatted Listing after deletion of a unit

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
GGEORPB69RS	9010	A	5	99	6APPBS
	(notice unit has been deleted)				
GGEORPB69RS	9030	A	5	99	6CADMINISTRATIVE ORGANIZATION

Example 4: Deleting a reference

Formatted Listing before deletion of the reference

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
GGEORPB69RS	9030	A	5	99	6CADMINISTRATIVE ORGANIZATION
OSTPLMC69MO	0010	A	1		ENTRY DATE 03/69
OSTPLMC69MO	0030	A	2		–#OSTPLMC69MO
•	•	•	•		•
•	•	•	•		•
•	•	•	•		•
OSTPLMC69MO	9010	A	5	99	6AREGIONAL PLANNING
PRESTCP69RL	0010	A	1		ENTRY DATE 03/69

Form of input (for deletion)

OSTPLMC69MO		D	1		

Formatted Listing after deletion of a unit

Document #	*Unit #*	*Code*	*Type*	*Elem #*	*Text*
GGEORPB69RS	9030	A	5	99	6CADMINISTRATIVE ORGANIZATION
	(notice the reference has been deleted)				
PRESTCP69RL	0010	A	1		ENTRY DATE 03/69

Document Master File Revisions

These procedures are concerned with the additions, deletions and changes to records already on the Document Master File. Such changes as are made are the result of policy changes in the form of entry or the need to add data to records already on the file. In the course of updating the Document Master File, new errors may occur. These become evident as a result of the editing programs operating during the master file update operation. Judgement must be exercised in making only those changes that are necessary for the integrity of the file or the products of the system.

The master file update procedures provide the following computer editing:

Verification of the sequence of the input records;

Verification of the indexing terminology;

Editing for correct codes and sequences in the free text segments;

Checking to be sure that systems restrictions as to size of the record, number of descriptors assigned, etc. have not been violated.

For any detected errors, an error schedule explaining the nature of the particular error is provided for the document analyst. In some cases the error is corrected by re-entry of that particular unit of information as an addition to the master file record. In other cases, the correction is handled by deleting information from the file. If a decision is made to ignore the errors, no further manipulation of the record is required. Since the units responsible for the error message have not been added to the Document Master File, this course of action does not result in a file containing incorrect data.

To prepare additions or deletions to the Document Master File, the analyst must reference the Formatted Listing of the documents being revised for document number and unit number. In some cases rank number is needed and this is obtained from the printout of the Document Master File record. A special worksheet has been prepared for Document Master File revisions. The analyst works with the error messages, error schedule, Formatted Listing and new master file update. All requisite changes are entered on a Master File Edit Worksheet. A separate worksheet is prepared for each document requiring changes.

In the case of additions the worksheet must contain:

The document number of the record to which the addition is being made;

The unit number for the data that is being added;

The entry code 'A' (for addition) in column 19;

Coding for the type of data being entered – '2' for descriptors, '3' for subdescriptors, and '5' for free text segment;

For a free text segment, the element number;

For the addition of a subdescriptor to a descriptor already on the file, the rank number of that descriptor (a number assigned by the program to sequence the descriptors within a reference). The rank number is found on the Document Master File printout for that document;

For descriptor and subdescriptor additions, the requisite pre-code: '*', '#', 'N', etc.;

The data being added.

Example 1: Adding a descriptor and subdescriptor to a reference

Worksheet entry

```
1                14 15   18 19 20 23                                  46 59       70
OSTPLMC69M0         0560    A  2  *INTERLOCAL COOPERATION               MUNICIPAL
```

Revised reference - since the last existing descriptor was Rank 013, the addition reads:

RANK014 *INTERLOCAL COOPERATION/001/QUAL001 MUNICIPAL

Example 2: Adding a subdescriptor to a descriptor already in the reference

Original reference

RANK012 *STATISTICS/000/

Worksheet entry

```
1                14 15      18 19 20 21          34 35       46
BASPOPA69FPES       0600      A  3  12N STATIS      BY-CITY
```

Revised reference — the subdescriptor will have been added

RANK012 *STATISTICS/001/QUAL001 BY—CITY/

Example 3: Adding free text to the reference

Worksheet entry

```
1                14 15 18    19 20 21-22 23
PRESTCP69RL         4040     A  5  56    15 Founder St., Reston, Va.
```

Revised reference — the text will have been added

```
A5  PRESTCP69RL     ENTRY 07/69                    SEGMENT NO.056
    15 FOUNDER ST., RESTON VA.
```

Deletions to the record occur when some of the data in the master file record is to be eliminated. As in the case of additions, delete instructions must be accompanied by sufficient data to guarantee that the unwanted data will be located and eliminated. In the case of deletions, the worksheet must contain:

The document number;

The unit number of the data being deleted;

The delete code 'D' in column 19:

The rank number for descriptors or subdescriptors that are to be deleted;

In the case of free text segments:

The segment number appropriate to that data;

The first six characters of the segment to be deleted; except for subject headings, the entire segment must be deleted and this procedure insures deletion of the entire segment;

Only an individual subject heading may be deleted from a free text segment;

When deleting an entire document it is not necessary to delete the document unit by unit. Deleting the heading unit removes the entire document.

Example 4: Deleting a descriptor from a reference

Original Document Master File entry

```
RANK008     *RENEWAL  /000/RANK009  *GLOSSARIES/000/
```

Worksheet entry

```
1                   14    15    18    19    20    21                  34
 LINDIMAMP68HR          0530        D    2     08    *RENEWAL
```

Revised Document Master File entry

```
RANK008     *GLOSSARIES  /000/
```

Example 5: Deleting a subdescriptor from a descriptor in a reference

Original Document Master File entry

```
RANK012     *FINANCIAL PROGRAMMING  /001/QUAL 001    FORECASTS /
```

Worksheet entry

```
1                  14   15   18   19   20   21          34   35          46
 KNEWCPCR68RF           0750      D    3   12N FINANC     FORECASTS (sub-
                                                            descriptor being
                                                            deleted)
```

Revised Document Master File entry

```
RANK012     *FINANCIAL PROGRAMMING  /000/
```

Example 6: Deleting free text from a reference

Original Document Master File entry

```
TEXT A5  KNEWYHD. .69LE  ENTRY  07/69  SEGMENT NO. 009
         LOWER EAST SIDE,  NEW YORK (Segment being deleted)
```

Worksheet entry

```
1                  14   15   18   19   20   21-22   23        27
 KNEWYHD. .69LE         0530      D    5    09      LOWER
```

Example 7: Deleting a subject heading

Original Document Master File entry

```
TEXT      A5   DNATCUP68BA102    ENTRY 07/69    SEGMENT NO.099
                6ACOOPERATIVES
                6BTURNKEY
                6CMORTGAGE FINANCING
                6DGOALS
```

Worksheet entry

```
1                   14   15   18   19   20   21-22   23
 DNATCUP68BA102          9010      D    5    99      6ACOOPERATIVES
```

Revised Document Master File entry

```
TEXT      A5   DNATCUP68BA102    ENTRY 07/69    SEGMENT NO.099
                6BTURNKEY
                6CMORTGAGE FINANCING
                6DGOALS
```

Example 8: Deleting an entire reference

Worksheet entry

```
1                   14   15   18   19   20
 GGEORPB69RS             0010      D    1
```

Change instructions are concerned with data which are on the record. A change is processed as the deletion of the existing data and the addition of new data. Delete instructions, as outlined, must precede change instructions. In the case of changes, the worksheet must contain the deletion instruction followed immediately by the addition instruction for that unit.

All changes are treated as additions with the following differences:

Changes for data deleted uses the next available unit number;

Changes contain an 'A' in column 19;

Changes in free text use the appropriate segment number and must include the complete free text data, except for subject headings;

Changes of document number requires deletion and reentry of the entire document.

Example 9: Changing document number

Original Document Master File entry

```
ADDITION    AD BASPO69     ENTRY 05/69    SEGMENT NO.000
                RANK001-=BASPOPA69  /000/RANK002 =SPECIAL
                 REPORTS   /000/RANK003 =DATES /000/
```

Worksheet entries

```
1                14   15    18    19   20
 BASPOPA69            0010        D    1
```

The document reference will be reentered under the new number.. There will be no data for the reference BASPO69.

Revised Document Master File entry

```
ADDITION    AD   BASPOPA69MRCH     ENTRY  07/69   SEGMENT NO.000
                RANK001- =BASPOPA69MRCH     /000/RANK002
                 =SPECIAL REPORTS  /000/
```

Example 10: Changing free text segment

Original Document Master File entry

```
TEXT    A5  DNATCUP68BA    ENTRY 06/69    SEGMENT NO.021
                CPO. 1968.504P.
```

Worksheet entry

1	14	15	18	19	20	21-22	23
DNATCUP68BA		3010		D	5	21	CPO. 1
DNATCUP68BA		3015		A	5	21	GPO. 1968. 504P.

Revised Document Master File entry

TEXT A5 DNATCUP68BA ENTRY 07/69 SEGMENT NO.021
GPO. 1968. 504P.

Program Control Card — Document Worksheet

Column	*Contents*
1	'1'
2-14	'A's
15	' '
16-18	'&'s
19	'/'
20	' '
21-22	'&'s
23	'1'
24-34	'A's
35	'1'
36-58	'A's
59	'1'
60-80	'A's

Search Data Entry

Basic Search Worksheet

Completing a search worksheet requires:

Organizing the terms into the proper logical expression;

Selecting the various options for the retrieval bibliography; and

Entering the various information into the appropriate sections of the worksheet.

Requestor Information

SEARCH REQUESTED BY _____ Name of the individual making the request.
COMPANY _____ Company for which he works.
INQUIRY NUMBER _____ Number of the specific request.

Searcher _____
Date_____ Prelim_____ Final_____

Name of individual running the search and the dates on which it was run.

This information is not entered as part of the machine-readable input for the search.

Title Section

					Number of Requestor				R C	Number of request within requestor							DATE Y Y M M					TITLE	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26		80
		0	1	T																							
		0	2	T																							

The Title Section consists of either one or two units. The purpose of this section is to provide identifying information about the requestor and/ or his request to be printed on the retrieval output.

The request number is entered in columns 1-2. This is an accession number for the request within the input batch. The first title card contains specific fields in columns 6-24. The following are suggested usages of these fields which could be changed if the user so desired:

cols 6-9: requestor identification;
col 11: search of current or new material 'C' or retrospective search on the entire file 'R'
cols 12-16: question number within the individual requestor;
cols 18-19: must be left blank because the accession number, required by the system in columns 1-2 of every card of the search input, will be reproduced here;
cols 21-24: date of the search in form of yymm, where yy = year and mm = month;
cols 26-80: name, address, telephone, etc.

The second title card need not be used. However, if additional information is desired, it is entered in columns 6-80. Columns 1-2 will contain the request number.

Both units for the Title Section have been precoded for columns 3-5. Columns 3-4 contain the unit number within the request. Column 5 contains the unit code.

Format Section.

									B D M		Y N			F D			A N	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			
		0	3	F	/	B	,	/		,	/	D	,		A			

The purpose of the Format Section is to select the contents for the retrieval output. The analyst must specify whether descriptors are to be included and to what extent the bibliographic material for each reference is to be presented.

Columns 1-2 contain the request number. Columsn 3-4 have been precoded for unit number and column 5 for unit code.

The CFS system allows three types of searches: a Descriptor or Boolean Search, 'B'; a

search for specific document references, Document Search, 'D'; and a descriptor search for a list of specific document references, Mixed Search, 'M'. The proper search type must be entered in column 7. Since almost all of URBANDOC's searches were Descriptor or Boolean Searches, column 7 has been precoded for this search type. If one of the other types of search is to be performed, the analyst may override the precoding.

In formatting the retrieval output, descriptors can be included by specifying 'Y' in column 10. The absence of the descriptors can be handled by an 'N' in column 10.

The selection of bibliographic information starts in column 16. To include all the bibliographic information, column 16 should contain an 'A'. For no bibliographic information, column 16 should contain an 'N'. The user may specify the selection of specific elements for a particular request by listing the element numbers as three digit numbers separated by commas, i.e., '001,002,011,099'. A maximum of ten elements may be selected.

Request Section.

1	2	3	4	5	6	80
		0	4	R		

The Request Section expresses the search request as a combination of terms in specified *AND, OR* or *AND NOT* relationships. The Request Section for a basic search differs from an expanded search in the following ways:

The terms and the subdescriptors appear in the logical expression, (Boolean expression);

The user must sequence the terms in order of lowest frequency of usage; and

The user must expand the *AND, OR* and *AND NOT* relationships in the strings of the search term.

The form for the argument (or text) of the Request Section depends on the type of search (as indicated in the Format Section):

The request argument for a Boolean or Descriptor Search consists of the Boolean expression enclosed within a single set of parentheses;

The argument for a Document Search consists of the list of document numbers separated by commas with no punctuation after the last document number;

The argument for a Mixed Search is a combination of the Document Search and a Boolean Search. The left parenthesis at the beginning of the Boolean expression marks the end of the document list and the beginning of the Boolean expression.

The request number must be entered in columns 1-2 of every unit in the Request Section.

Columns 3-4 have been precoded for unit number and column 5 for unit code. There may be a maximum of thirty request units.

The request argument (or text of the request) is entered beginning in column 6. In writing the request, the following conditions must be met:

Pre-code must be included for descriptors;

All descriptors must agree with the form on the Inverted File, longer descriptors having been truncated to twenty-three characters;

When using the root of a document number as a descriptor, the '#' must be included as the first character of the document number in descriptor form;

Subdescriptors longer than twelve characters must be truncated after the twelfth character;

All Boolean operators must be the ones specified by the CFS system;

All operators and terms must be separated by a blank space and two adjacent operators are not permitted;

All Boolean expressions must be enclosed by a single set of parentheses, with an equal number of left and right parentheses;

A word may be split between two units only if it is continued through column 80 and resumes in column 6 of the next unit.

Expanded Search Worksheet

Completing an expanded search worksheet requires the same fundamental steps as completing a Basic Search Worksheet:

Organizing the terms into the proper logical expression;

Selecting the various options for the retrieval bibliography; and

Entering the various information into the appropriate sections of the worksheet.

Requestor Information. The Requestor Information for the Expanded Search Worksheet is identical to the Requestor Information for the Basic Search Worksheet. (For detailed instructions see Basic Search Worksheet.)

Title Section. The Title Section for the Expanded Search Worksheet is identical to the Title Section for the Basic Search Worksheet. (For detailed instructions see Basic Search Worksheet.)

BASIC SEARCH WORKSHEET

Searcher _____
Date _____ Prelim ____ Final ____

SEARCH REQUESTED BY _____
COMPANY _____ INQUIRY NUMBER _____

Cols. 1 & 2...Request No. On Request Card (R in col. 5) equation must start in col. 6.

| | | |

Format Section.

						B D M			Y N			F D			A N	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
		0	3	F	/	B	,	/		,	/	D	,		A	

Although the Format Section on both search worksheets appear identical, the analyst is offered fewer options on the expanded search worksheet, namely on the search type and the selection of bibliographic information.

Columns 1-2 contain the request number. Columns 3-4 have been precoded for unit number and column 5 for unit code.

Whereas the basic CFS system allows for three types of searches, the expanded search can operate only for a Descriptor or Boolean Search. If either of the other searches must be performed, the Basic Search Worksheet and basic search procedures must be used.

In formatting the *Retrieval Report*, descriptors can be included by specifying a 'Y' in column 10 of the Format Section. The absence of descriptors can be handled by an 'N' in column 10.

The selection of bibliographic information starts in column 16. To print all the bibliographic information, column 16 should contain an 'A'. For no bibliographic information, column 16 should contain an 'N'. The user may not specify the selection of specific elements for a particular request.

Term Table Section.

Request Terms — N=not equal, E=equal

1	2	3	4	5	6 Descriptor Table 41	42	43 Subdescriptor 53
		0	5	A			
		0	5	B			

The Term Table Section provides a table of correspondence between each unique term or form of a term to the code in the logical expression. Each table entry associates a term with a letter of the alphabet or a numerical digit. There may be a maximum of thirty-five entries in this section.

Each term in the request must be listed in the table. A term with no subdescriptors is entered into the table once. For a term with more than one subdescriptor, the term must be entered into the table once for each subdescriptor. The presence or absence of the subdescriptor also affects the term table. If both the presence and the absence of the same descriptor and subdescriptor is desired (admittedly an unusual combination for one request), then the term must be entered once for presence and once for absence. The following examples illustrate the number of entries required:

Descriptor A and Descriptor B. One table entry for Descriptor A, another for Descriptor B.

Descriptor A modified by B or C and Descriptor D. One table entry for Descriptor A modified by B, one for Descriptor A modified by C, and a final entry for Descriptor D.

Descriptor A modified by B and Descriptor C or Descriptor A not modified by B and Descriptor D. One table entry for Descriptor A modified by B, another for Descriptor C, a third for Descriptor A not modified by B, and a final entry for Descriptor D.

Should the number of entries exceed thirty-five, the analyst must divide the request into more than one question.

The analyst enters request number in columns 1-2 of the term table unit. Columns 3-4 have been precoded for unit number and column 5 for unit code.

When completing the term table entry; the following conventions must be met:

The descriptor must start in column 6;

All descriptors must agree with the form on the Inverted File, longer descriptors having been truncated to twenty-three characters;

Pre-code is not included for descriptors;

When using the root of a document number as a descriptor, the '#' must be included as the first character of the document number in descriptor form;

Each unique combination of a descriptor and subdescriptor requires an entry in the table;

The presence 'E' or absence 'N' of a subdescriptor must be specified in column 42;

The subdescriptor must start in column 43;

Subdescriptors longer than twelve characters must be truncated after the twelfth character; and

All symbols used in the table must be used in the logical expression.

Logical Expression Section.

1	2	3	4	5	6	80
		0	0	R		

The Logical Expression Section is designed to express the search request as a combination

of terms in specified *AND, OR or AND NOT* relationships. The logical expression differs from the expression of the request in the basic search in the following ways:

The code from the term table entry corresponding to the term is used in place of the term itself; and

The analyst may use parentheses to expand the *AND, OR* or *AND NOT* relationships in the request.

The use of the expanded search is of assistance to the analyst since the program will sequence the terms in the requests in order of lowest frequency of usage.

The request number is entered in columns 1-2 of the unit. Columns 3-4 have been precoded for unit number and column 5 for unit code.

When forming the logical expression, the following conventions must be met:

The expression starts in column 6;

All symbols and operators must be adjacent with no blank spaces within the expression;

Parentheses are to be used for anything that must be expanded by the program, but there must be an equal number of left and right parentheses;

All Boolean operators must be the ones specified by the CFS system;

Columns 79-80 must always be blank; and

The expression must fit in seventy-three characters.

Program Control Card – Search Worksheets

Column	*Contents*
1	'Ø'
2	'+'
3	' '
4	'+'
5	'1'
6-10	'A's
11	'1'
12-20	'A's
21	'1'
22-25	'A's
26	'1'
27-41	'A's
42	'1'
43-80	'A's

EXPANDED SEARCH WORKSHEET

Searcher _____

Date _____ Prelim ____ Final ____

SEARCH REQUESTED BY _____

COMPANY _____ INQUIRY NUMBER _____

Cols. 1 & 2...Request No. On Request Card (R in col. 5) equation *must* start in col. 6.

1	2	3	4	5	6	80
		0	0	R		

					Number of Requestor					R C	Number of Request within Requestor									DATE Y Y	M M			TITLE	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	80
		0	1	T																						
		0	2	T																						

						B D M				Y N		F D		A N		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
		0	3	F	/	B	,	/		,	/	D	,	/	A	

REQUEST TERMS ← N = not equal, E = equal

1	2	3	4	5	6	DESCRIPTOR TABLE 41	42	43 SUBDESCRIPTOR 53
		0	5	A				
		0	5	B				
		0	5	C				
		0	5	D				
		0	5	E				
		0	5	F				
		0	5	G				
		0	5	H				
		0	5	I				
		0	5	J				
		0	5	K				
		0	5	L				
		0	5	M				
		0	5	N				
		0	5	O				
		0	5	P				

☐ Check this box if more descriptors on continuation page.

Figure 13

PROCESSING CYCLES

During systems analysis (design through implementation), the computer programs, either existing or under development, were categorized into groups according to basic function. Five modules (Thesaurus, Pre-edit, File Maintenance, Search and Publications) evolved in which each module contained all the programs concerned with achieving a particular goal.

When studying the project activities for the Operational Data Analysis, it was discovered that the superstructure imposed on the programs was applicable primarily for systems analysis and programming. It was not applicable to an analysis of production because in most cases a process or a product did not correspond directly to a specific programming module. In some instances only a part of a programming module was used. In others, several modules, or parts of modules, were involved.

Since the previous discussions have been oriented towards systems analysis and programming, the programs were organized by module. The following discussions are oriented toward the operation of the system in terms of processing cycles.

Editing and Validation Cycle

The Editing and Validation Cycle is aimed at producing error-free document input which leads, hopefully, to error-free products. Some errors can be detected by computer; others can be detected only through proof-reading. The Editing and Validation Cycle uses the Pre-edit, File Maintenance and Publications Modules. (See Figure 14.)

The Pre-edit Module is responsible for computer validation of the sequence of the input, the formats and codes and the length of the content analysis terms. It is also responsible for the production of a proof-reading sheet and a master sheet for error correction.

The File Maintenance Module is responsible for the computer validation of the content analysis and subject heading terms against the Inverted File. It also checks the validity of combinations of information types within a reference.

The Publications Module is responsible for producing a sample copy of the *Input Index*. This copy allows the document analyst to check for spelling errors within the bibliographic information, coding errors in the element numbers, consistency of form within each bibliographic element, and the length of the pages for camera-ready copy.

The Editing and Validation Cycle is subdivided into three phases:

PHASE I. Preparation of the Data Base
1. Card-to-tape of the Document Input (E0010)
2. Sort the Document Input File (E0015)
3. Expand the Document Input File (E0020)
4. Edit the Document Input File (E0030)

PHASE II. File Maintenance and CFS Error Detection
1. Format the Document Input File (MAINT1)

2. Sort the descriptors, subdescriptors and subject headings (SORTM1)
3. Validate the descriptors, subdescriptors and subject headings (MAINT2)
4. Sort the error units (SORTM2)
5. Perform the CFS edit (MAINT3)

PHASE III. Create the sample *Input Index*

1. Create the Publication File and the optional Content Analysis File (P0010)
2. Sort the Content Analysis File (P0015)
3. Sort the Publication File (P0025)
4. Create the Subject Subset File (P0040)
5. Sort the Subject Subset File (P0045)
6. Print the Major Subject Listing (P0050)
7. Print the Main Document Listing (P0050)
8. Sort the Publication Index File (P0055)
9. Print the *Input Index* sections (P0060).

Input Processing Cycle

Once the Editing and Validation Cycle has been completed, there exists a body of changes to the Document Input File. Some changes are the addition of entirely new information while others are the correction of existing errors detected during Editing and Validation (both human and computer). The Input Processing Cycle updates the original input with these changes. The resulting file is checked once more for errors within the revision input or invalid combinations of information types as a result of the revision processing.

The Input Processing Cycle uses the Pre-edit and File Maintenance Modules. A portion of the Pre-edit Module updates the original input with the revisions. The File Maintenance Module is used for the final computer validation of sequence, coding, and combination of input types. (Note: If the user so desires, it is possible to bypass this final validation. However, URBANDOC experience indicates that much consideration should be given before doing so. See Chapter IV, specifically the abstract for MAINT3.) The final task of the Input Processing Cycle produces synchronized input for the creation of searchable files and the *Input Index*. (See Figure 15.)

The Input Processing Cycle is subdivided into three phases:

PHASE I. Updating the Thesaurus File

1. Pre-list the Thesaurus Input (XMAIN0)
2. Format the Thesaurus Input (XMAIN1)
3. Sort the Thesaurus Input File (SORTX1)
4. Update the Thesaurus File (XMAIN2)
5. Print the Thesaurus File (XMAIN3)

PHASE II. Preparation of the Document Input File

1. Card-to-tape of the Document Revision File (E0010)

EDITING AND VALIDATION CYCLE

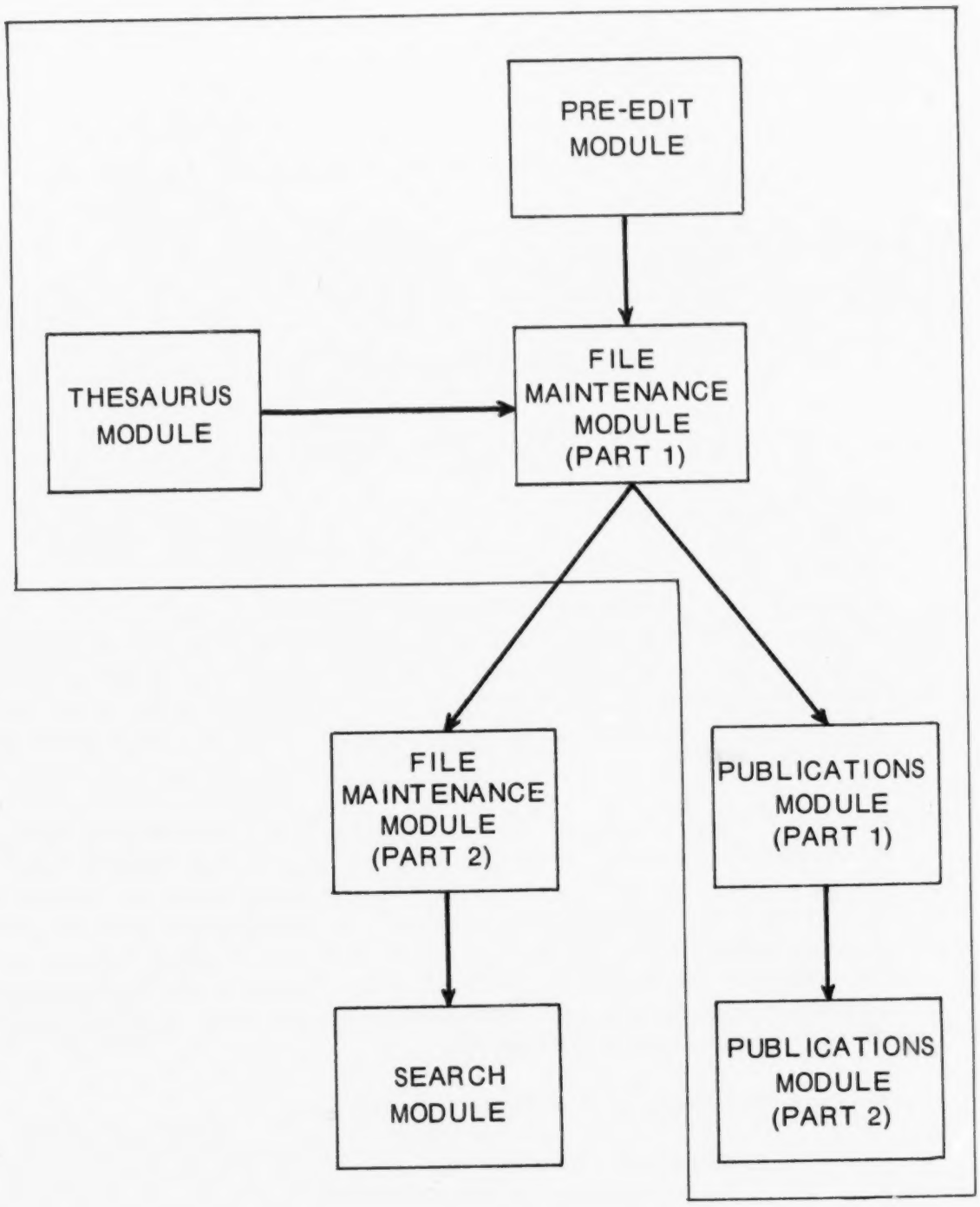


Figure 14

INPUT PROCESSING CYCLE

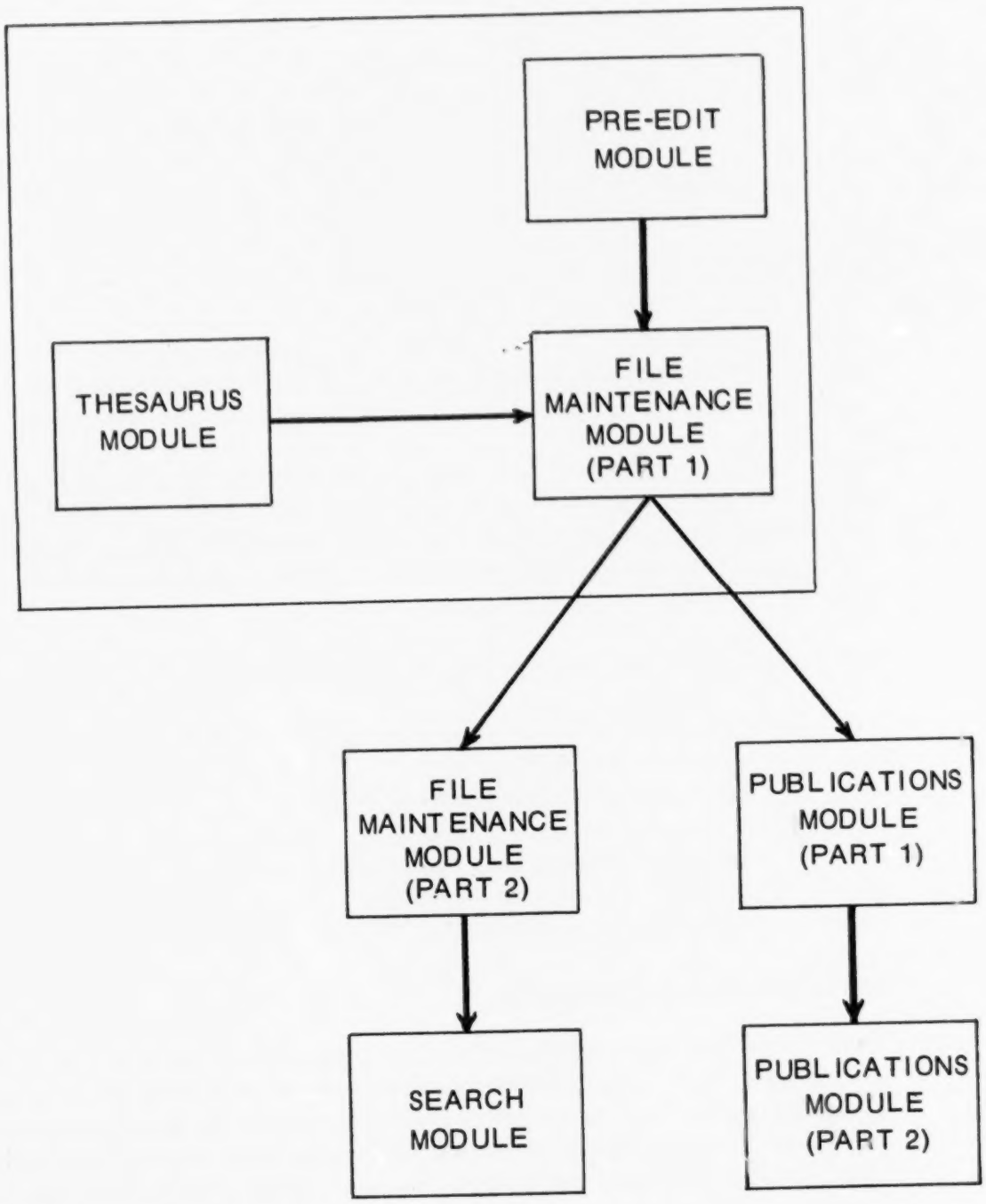


Figure 15

2. Sort the Document Revision File (E0015)
3. Update the Document Input File (E0040)

PHASE III. CFS File Maintenance Procedures
1. Format the Document Input File (MAINT1)
2. Sort the descriptors, subdescriptors and subject headings (SORTM1)
3. Validate the descriptors, subdescriptors and subject headings (MAINT2)
4. Sort the error units (SORTM2)
5. Perform the CFS edit (MAINT3)

Input Index File Maintenance Cycle

Once the synchronized input for publications and search has been produced, each product is created separately. For purposes of cost analysis, creation of the *Input Index* is split into file maintenance of the data base and production of the camera-ready copy.

Input Index File Maintenance consists of selecting the input, formatting it into the appropriate auxiliary publications files and sequencing each file according to the report it will produce, in this case the tape files for the Main Document Listing, Major Subject Listing and indexes to the documents. *Input Index* File Maintenance uses the formatting programs of the Publications Module. (See Figure 16.)

The *Input Index* File Maintenance procedures are as follows:

1. Create the Publication File and the optional Content Analysis File (P0010)
2. Sort the Content Analysis File (P0015)
3. Sort the Publication File (P0025)
4. Create the Subject Subset File (P0040)
5. Sort the Subject Subset File (P0045)
6. Sort the Publication Index File (P0055).

Input Index Production Cycle

Input Index File Maintenance covers the processing up through the creation of camera-ready copy. *Input Index* Production covers the production of the camera-ready copy. This has been kept separate from the maintenance of the tape files since it could be printed by another organization or another method such as photocomposition. Any change in the procedures for creating this copy would undoubtedly affect both the operations and cost for this aspect of a documentation effort. As such it should be reflected in the discussions on operations.

The current *Input Index* Production Cycle calls for the printing of the Major Subject Listing, the Main Document Listing and such indexes to the reference citations as author and title indexes. *Input Index* Production uses the report-writing programs of the Publications Module, including some formatting programs that produce final copy of the double column indexes. (See Figure 17.)

INPUT INDEX FILE MAINTENANCE CYCLE

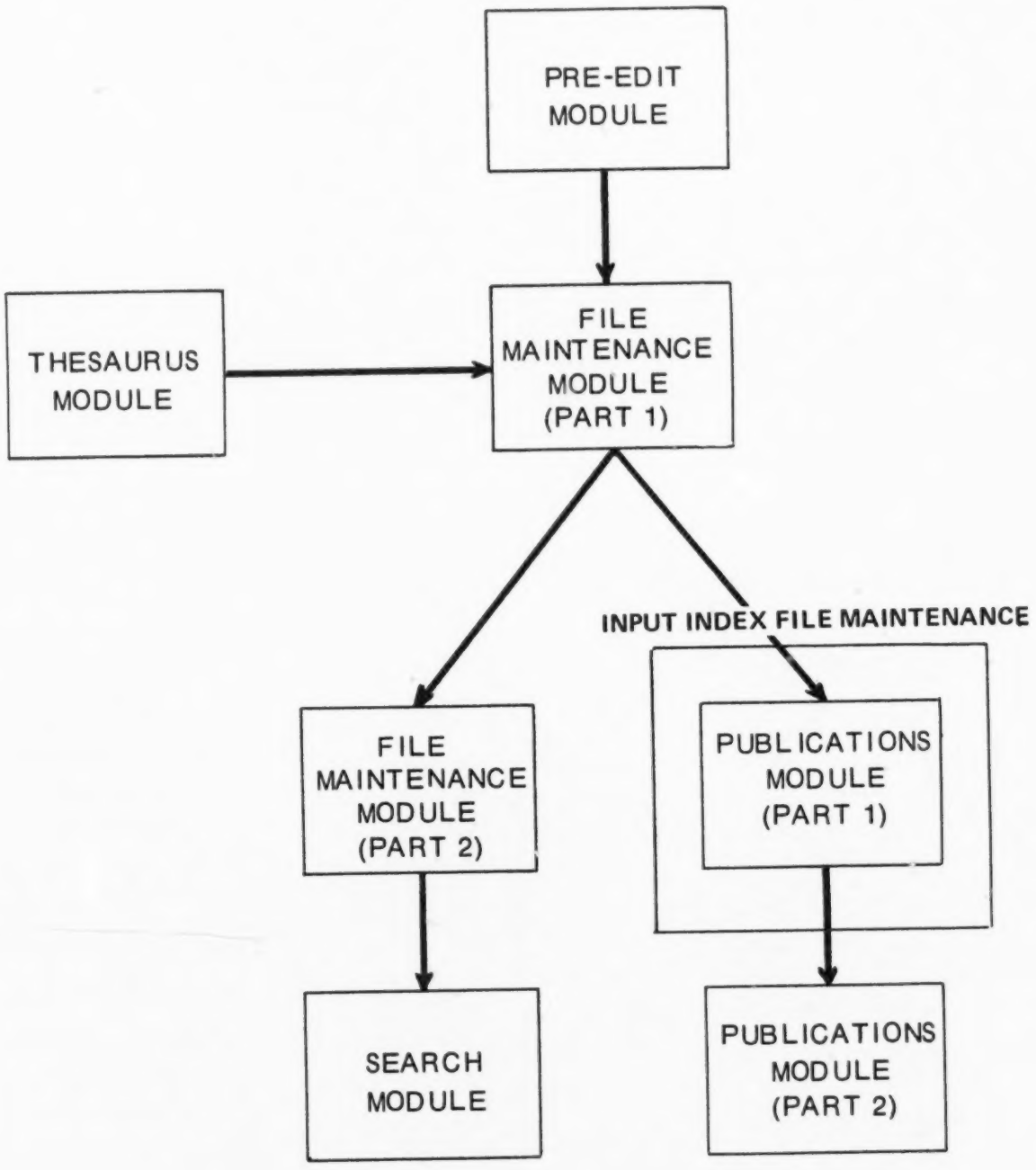


Figure 16

INPUT INDEX PRODUCTION CYCLE

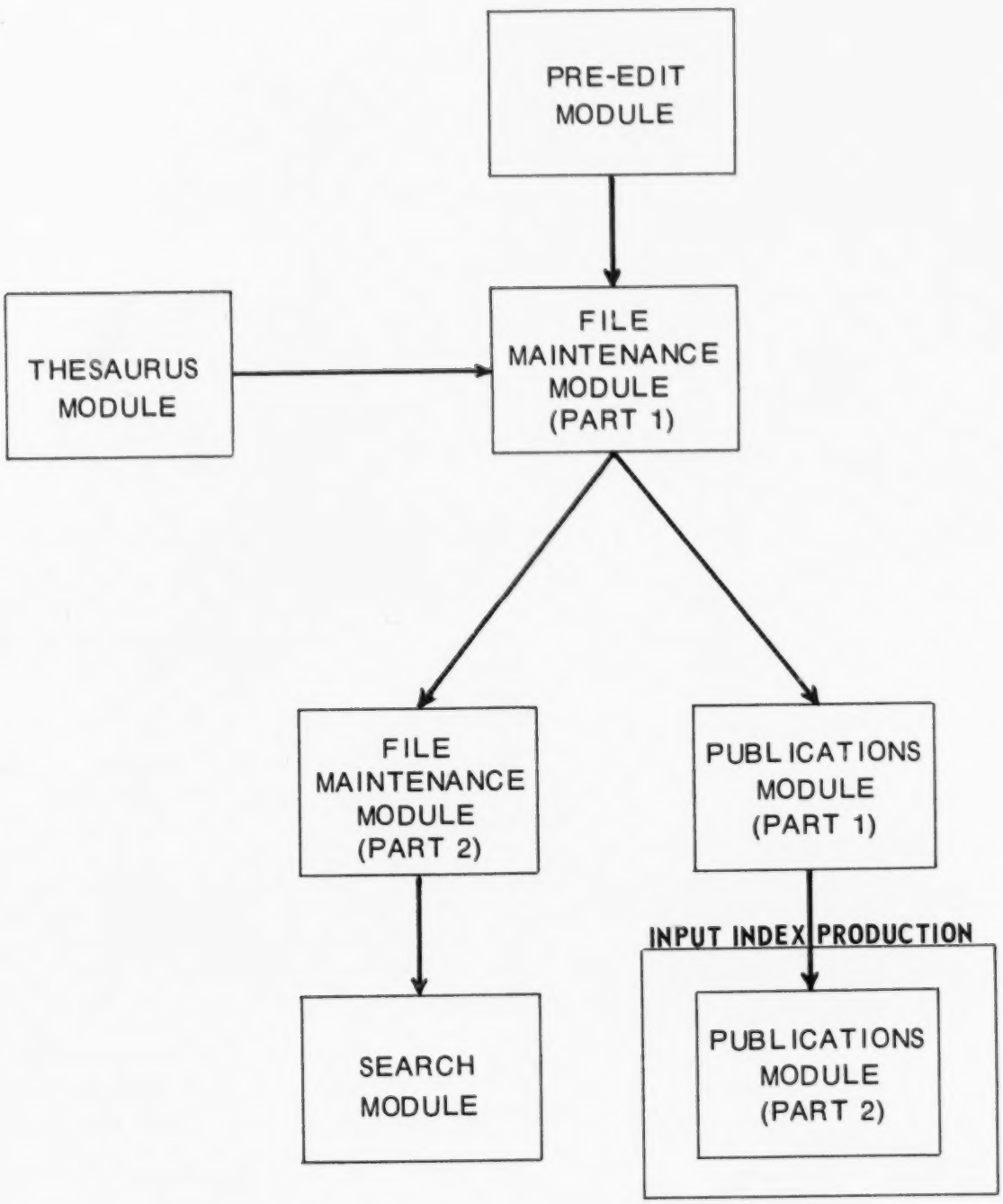


Figure 17

The *Input Index* Production procedures are as follows:

1. Print the Major Subject Listing (P0050)
2. Print the Main Document Listing (P0050)
3. Print the single column indexes and create the double column indexes (P0060)
4. Sort the double column index sections (P0065)
5. Print the double column indexes (P0070).

Retrieval Report File Maintenance Cycle

Retrieval Report File Maintenance uses the portion of the File Maintenance Module devoted to the support of the searchable files. The output of the Input Processing Cycle updates the Document Master File and the Inverted File. Listings of the processed transactions and updated files and analyses of the Document Master File and the Inverted Files are made available. At the completion of this cycle, there exists a set of machine-readable files for computer searching and printed tools for manual searching. (See Figure 18.)

The *Retrieval Report* File Maintenance procedures are as follows:

1. Update the Document Master File (MAINT4)
2. Analyze the Document Master File (MAINT8)
3. List the transactions and create the Descriptor Input File (MAINT6)
4. Sort the Descriptor Input File (SORTD1)
5. Update the Inverted File (DMAIN1)
6. Print the Inverted File Summary Listing (DMAIN2)
7. Print the Inverted File Detail Listing (DMAIN3).

Retrieval Report Production Cycle

Retrieval Report Production covers the computer processing of a request for information through the creation of a specialized bibliography by either basic or expanded search procedures. The basic search performs an inverted file search and a serial file search to produce the bibliography of retrieved references; term validation is not provided under this procedure. Additionally, optimum performance of the search cannot be guaranteed. The expanded search procedure is used to expand the search request format, verify that each term has been used in indexing and format the search request to assure that optimum processing will occur. (See Figure 19.)

The *Retrieval Report* Production Cycle is subdivided into two phases:

PHASE I. Search Diagnostics
1. Expand the requests (SEP1)
2. Sort the request terms (SORT SEP1)
3. Match the request terms to the Inverted File (SEP2)
4. Sort the request terms (SORT SEP2)
5. Format the request to CFS specifications (SEP3)

RETRIEVAL REPORT FILE MAINTENANCE CYCLE

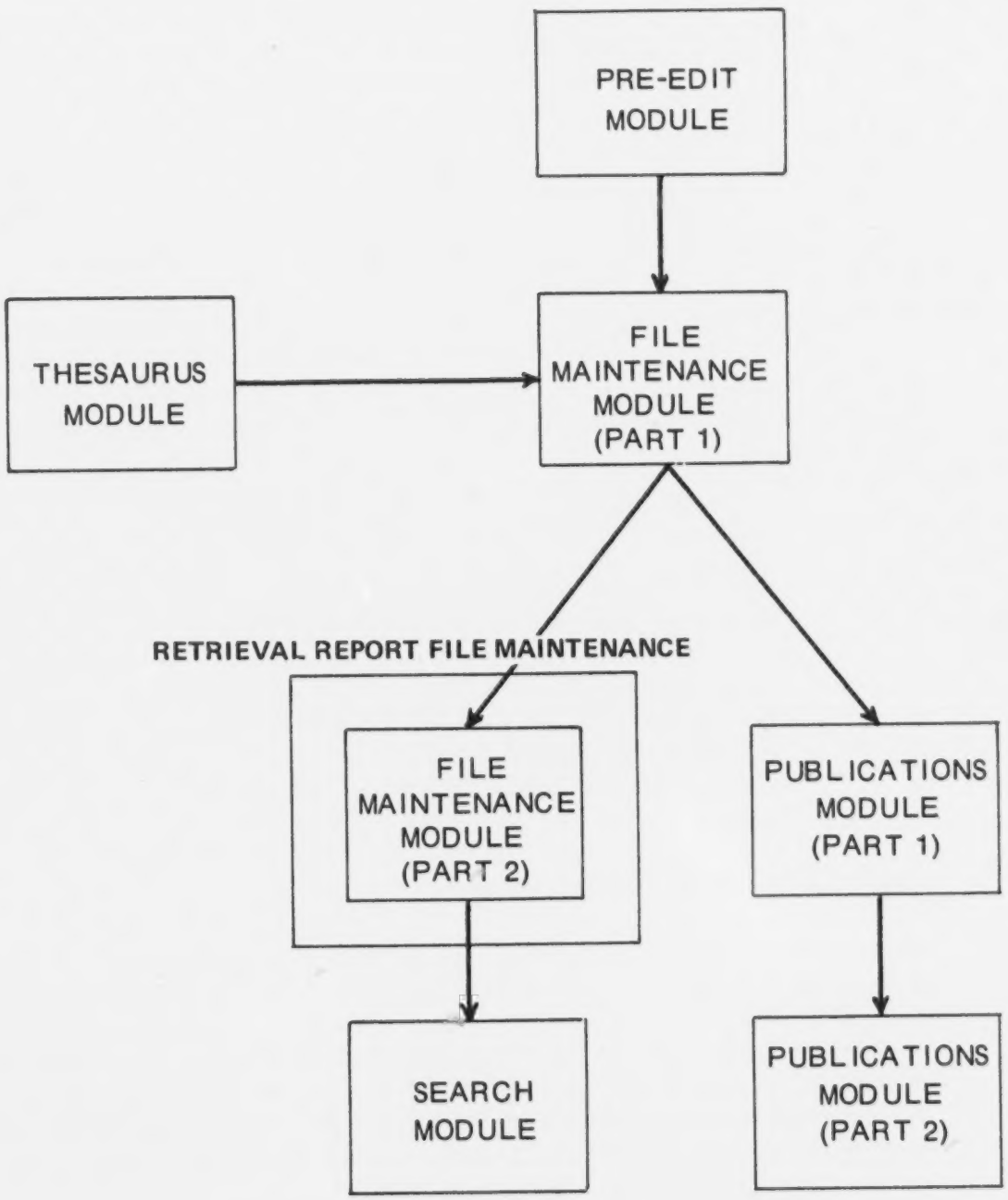


Figure 18

RETRIEVAL REPORT PRODUCTION CYCLE

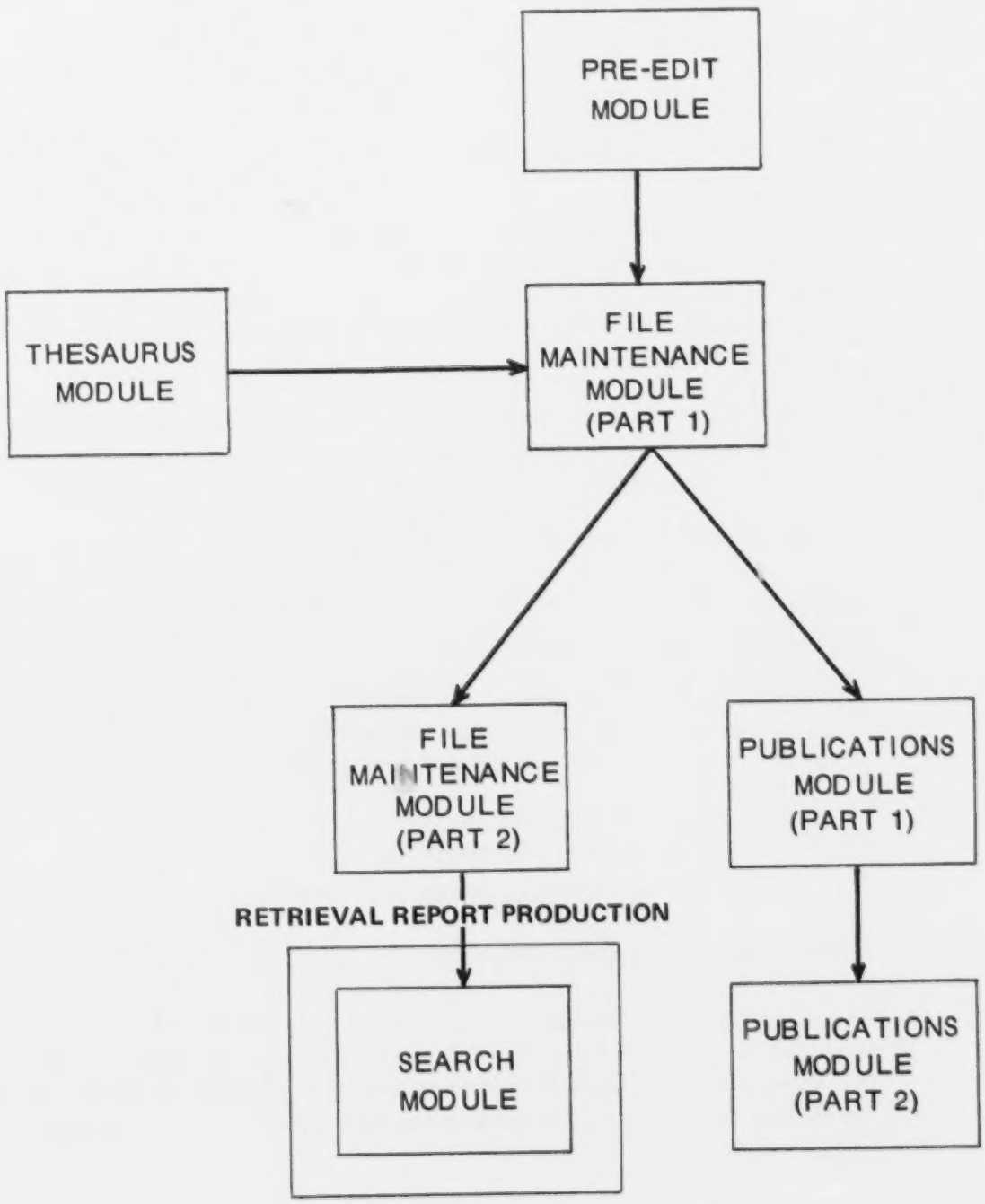


Figure 19

PHASE II. CFS Search

1. Edit the requests (PHASE1)
2. Inverted file search (PHASE2)
3. Boolean reduction merge (PHASE3)
4. Prepare for the serial file search (PHASE4)
5. Serial file search (PHASE5)
6. Print the *Retrieval Report* (PHASE6).

Miscellaneous Thesaurus Products

This processing cycle covers the performance of those runs concerned with the Thesaurus File that are performed on an on-demand basis and as such do not form part of the standard processing. Such miscellaneous products include an analysis of the Thesaurus File, printing the Permuted Thesaurus and creating supplements of new editions of the URBANDOC Thesaurus.

The operations discussion covers several of the products through the production of hard-copy output. In the event that creation of machine-readable files would be sufficient, the procedures may be terminated with the creation of the file.

There are four basic processing options:

Option I. Structural Analysis of the Thesaurus File (XMAIN4)

Option II. Print the Permuted Thesaurus File (XMAIN5)

Option III.. Create a Thesaurus Supplement
1. Create the Thesaurus Supplement File (XMAIN6)
2. Print the Thesaurus Supplement (XMAIN3)

Option IV. Publish a Thesaurus Edition
1. Print the Thesaurus (XMAIN3)
2. Print the Permuted Thesaurus (XMAIN5).

Miscellaneous Document Products

This cycle covers the performance of runs concerned with the Document Master File produced on an on-demand basis. As such, they do not form part of the standard processing of the file. Such miscellaneous products include an analysis of the Document Master File, subsetting the file, merging two separate files, and creating new book catalogs and indexes to the file.

The operations discussion covers several of the products through the creation of hard-copy output. In the event that machine-readable files are sufficient, the procedures can be terminated with the creation of the files.

There are five basic processing options:

Option I. Structural Analysis of the Document Master File (MAINT8)

Option II. Document Master File Merge
1. Merge two Document Master Files and create a combined Descriptor Input File (MAINT7)
2. Sort the Descriptor Input File (SORTD1)
3. Create an Inverted File (DMAIN1)
4. Print the Inverted File Summary Listing (DMAIN2)
5. Print the Inverted File Detail Listing (DMAIN3)

Option III. Create a Document Master Subset File (MAINT9)

Option IV. Update the Book Catalog (LIBPRT)

Option V. Create Indexes to the Document Master File
1. Create a Publication File (P0010)
2. Sort the Publication File (P0025)
3. Print the single column indexes and create the double column indexes (P0060)
4. Sort the double column indexes (P0065)
5. Print the double column indexes (P0070).

Miscellaneous Publications Products

This processing cycle covers the creation of special editions and cumulative issues of the *Input Index*. Special editions are usually created on an on-demand basis while cumulative issues may be scheduled for annual or semi-annual release. In either instance they do not warrant forming a part of the standard publications processing cycles.

The operations discussion covers the products through the creation of printed output. If the existence of the machine-readable files is sufficient, the procedures may be terminated with the creation of the machine-readable files.

There are two processing approaches that may be taken:

Option I. Create a special *Input Index* or a cumulative *Input Index*
(Note: This method of creating a cumulative index should be used when the descriptive analysis information has been significantly revised since the creation of the issues in which the references first appeared.)
1. Create a special Publication File (P0080)
2. Sort the Publication File (P0025)
3. Create the Subject Subset File (P0040)
4. Sort the Subject Subset File (P0045)
5. Sort the Publication Index File (P0055)
6. Print the Major Subject Listing (P0050)
7. Print the Main Document Listing (P0050)
8. Print the single column indexes and create the double column indexes (P0060)
9. Sort the double column indexes (P0065)
10. Print the double column indexes (P0070)

Option II. Create a cumulative *Input Index* (Note: This method should be used when the contents of the descriptive analysis information has not been significantly changed since the creation of the issues in which the references first appeared.)

1. Cumulate the Publication Files (P0025)
2. Cumulate the Subject Subset Files (P0055)
3. Cumulate the Publication Index Files (P0045)
4. Print the Major Subject Listing (P0050)
5. Print the Main Document Listing (P0050)
6. Print the single column indexes and create the double column indexes (P0060)
7. Sort the double column indexes (P0065)
8. Print the double column indexes (P0070).

OPERATING INSTRUCTIONS

(Note: A numeric zero for operations data will be represented by the symbol 'Ø'.)

Thesaurus Module

XMAIN0

Function: To provide a proof-listing of the Thesaurus Input and detect certain types of errors.

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program XMAIN0	N/R	Program Library
Date Card	N/R	Discard
Thesaurus Input	N/R	Off-line Storage

Report

Name	Paper Specs	Carriage Tape Specs
Pre-edit of Thesaurus Input	Standard	Standard

Operator Messages

Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	To produce subdescriptor listing
C	To bypass sequence check
D	No publications substitute

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
001	Wrong date card	Correct and resubmit. Press Start.
002	Operator should set sense switches	Press Start.
999	End of job	

Special Instructions

Standard Sense Switch Settings: A, C, D

Operations Data — Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-9	mm/dd/yy

XMAIN1

Function: To format the Thesaurus Input

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program XMAIN1	N/R	Program Library
CFS Date Card	N/R	Discard
Thesaurus Input	N/R	Off-line storage

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Thesaurus Input File	3	Scratch	No	SORTX1

Card Output

File Name	Output Stacker	Disposition
SORTX1 Control Card	8/2	SORTX1

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages

Sense Switch Settings
Program Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	Single unit terms

Loading Instructions

Card Load

Program Halt List

Refer to Operator Messages

Special Instructions

Standard Sense Switch Setting: A

Operations Data — CFS Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-7	mmddyy

SORTX1

Function: To sort the Thesaurus Input File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library

SORTX1 Control Card	N/R	Discard
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Thesaurus Input File	3	XMAIN1	No	Tape Library
Work Tapes	1, 2, 4	Scratch	No	
Sorted Thesaurus Input File		Scratch	No	XMAIN2

(The Sorted Thesaurus Input File will be created on either Work Tapes 1 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors or Wrong Length Records	N/R	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data — SORTX1 Control Card

Field	Cols	Contents
Tape drive assignments	1-2	'31'
	4-5	'24'
Number of input reels	7-8	'Ø1'
Record length	9-12	'1299'
Input blocking	13-15	'ØØ1'
Output blocking	16-18	'ØØ1'
Tape reading mode	19-20	'P1'
Header/trailer labels	21-25	'11111'
Core storage size	27	'4'
Number of control fields	28-29	'Ø1'

Number of control characters	30-32	'040'
Control field 1	33-36	'0016'
Size of field 1	37-39	'040'
Record format	57	'1'
Number of characters in smallest record	62-65	'0146'
Identification	76-80	'XMAIN1'

XMAIN2

Function: To create and update the Thesaurus File

Card Input

File Name	Output Stacker	Disposition
Program XMAIN2	N/R	Program LIbrary

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Prior Thesaurus File	2	Tape Library	Yes	Tape Library
Sorted Thesaurus Input File	3	SORTX1	No	Tape Library
Updated Thesaurus File	4	Scratch	No	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages

Sense Switch Settings
Program Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card
C	Create file for first time

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
444	Tape write error on label for tape unit 4	Press Start to re-try
004	Tape write unit on drive 4	Press Start to try ten times more
999	End of job	

Special Instructions

Standard Sense Switch Settings: A

XMAIN3

Function: To print the subject section or the geographic section of the URBANDOC Thesaurus File

Card Input

File Name	Output Stacker	Disposition
Program XMAIN3	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Thesaurus File	2	XMAIN2	Yes	Tape Library
Work Tape	3	Scratch	No	Scratch

Report

Name	Paper Specs	Carriage Tape Specs
URBANDOC Thesaurus	Standard	Standard

Operator Messages

Sense Switch Settings
Progam Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card
B	Interrupt printing at end of page
C	Only geographic terms
D	Only subject terms
E	No publications substitutes
F	No dates
G	Punch new Permuted Thesaurus Input

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
002	Tape read error on tape unit 2	Press Start to try ten times more.
003	Tape write error on tape unit 3	Press Start to try ten times more.

Special Instructions

If the run should be interrupted after printing a page, put on Sense Switch B and save the tape on tape unit 3.

Standard Sense Switch Settings:
A, D, E, F for Subject Thesaurus
A, C, E, F for Geographic Thesaurus

For publishable copy of Thesaurus or a new Permuted Thesaurus, insert patches for XMAIN3 before last card of program deck. At the end of the run, be sure to remove the publication patches.

XMAIN4

Function: To print the Statistical Analysis of the Thesaurus File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program XMAIN4	N/R	Program Library
Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Thesaurus File	2	Tape Library	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Statistical Analysis	Standard	Standard

Operator Messages

Program Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
001	Date card missing or incorrect	Insert date card and press Start.
002	Tape read error	Press Start to try ten times more.
999	End of job	

Operations Data-Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-9	mm/dd/yy

XMAIN5

Function: To print the Permuted Thesaurus

Card Input

File Name	Output Stacker	Disposition
Program XMAIN5	N/R	Program Library
Title Card	N/R	Discard
Stop Words	N/R	Discard
Input Data	N/R	Off-line Storage

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Prior Permuted Thesaurus File	)	Tape Library	Yes	Tape Library
Work Tapes	) 1, 2, 3, 4	Scratch	No	Scratch
Updated Permuted Thesaurus File	)			Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Permuted Thesaurus	Standard	Standard

Operator Messages

Tape Mounting Instructions
Program Halts
Indication of Output Drive

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Program Halt List

Number (I-star)	Reason	Re-start
1617	Tape error routine	Job must be re-started.
1639	Tape error routine	Job must be re-started.
1785	Tape error routine	Job must be re-started.
2511	Stop word table filled	Re-start the job with fewer stop words
1503	Halt to allow merge of prior Permuted Thesaurus File	Mount the prior Permuted Thesaurus File on the drive indicated in the operator message and press Start.
1508	End of reel	Re-start the job with fewer entries in the Permuted Thesaurus File
1228	End of job	

Special Instructions

There is an option to mount a prior Permuted Thesaurus File. Operator message will tell the correct drive and continuation procedures.

Output tape will be indicated at the end of the run.

To produce several listings the operator need only follow the instructions at the end of the run.

Operations Data-Title Card

Field	Cols	Contents
Title	1-69	

Date	70-79	mm/dd/yy
Code	80	'*'

Operations Data-Stop Words

Field	Cols	Contents
Stop Word	1-20	
Card Code	78-80	'XXX'

XMAIN6

Function: To create the Thesaurus Supplement File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program XMAIN6	N/R	Program Library
Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Thesaurus File	1	Tape Library	Yes	Tape Library
Thesaurus Supplement File	2	Scratch	No	XMAIN3

Operator Messages

Program Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
111	No tape mark after header label	Must recreate Thesaurus File.
002	Tape read error	Press Start to try ten times more.
003	Tape write error	Press Start to try ten times more.
888	Tape write error for label on output tape	Re-start job.
999	End of job	

Operations Data-Date Card

Field	Cols	Contents
Code	1	'$'
Date	2-9	mm/dd/yy

Pre-edit Module

E0010

Function: To format the original and/or revision input to the requirements of the Pre-edit and File Maintenance specifications

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program E0010	N/R	Program Library
Lead Card	N/R	Discard
Document Input	N/R	Off-line Storage

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Input File	1	Scratch	No	E0015 or E0020

Report

Name	Paper Specs	Carriage Tape Specs
Listing of Sequence Errors	Standard	Standard

Operator Messages

Sequence Errors

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead card not '= 'LE0010'	Correct and resubmit. Press Start.
03	Option card not = 'DOC' or 'THS'	Correct and resubmit. Press Start.
99	End of job	

Special Instructions

If no sequence errors occur, a printed message will instruct the operator to go directly to E0020

Operations Data-Lead Card

Field	Cols	Contents
Code	1-6	'LE0010
Option	7-9	'DOC' if document Input; 'THS' if Thesaurus Input

E0015

Function: To sort the Document Input File into sequence by document number and unit number

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
E0015 Control Card	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Input File	1	E0010	Yes	Tape Library
Work Tapes	2, 3, 4	Scratch	No	
Sorted Document Input File		Scratch	No	E0020

(The Sorted Document Input File will be created on either Work Tapes 2 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors or Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save tape input

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter.

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-E0015 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'12'
	4- 5	'34'

Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ84'
Input blocking	13-15	'Ø1Ø'
Output blocking	16-18	'Ø1Ø'
Tape reading mode	19-20	'P1'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'Ø8Ø'
Control field 1	33-36	'ØØØ1'
Size of field 1	37-39	'Ø8Ø'
Identification	76-80	'E ØØ15'

E0020

Function: To expand the Document Input File and generate all the standard input units. An optional listing for proof-reading and error correction may be produced.

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program E0020	N/R	Program Library
Date Card	N/R	Discard
Lead Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Input File	1	E0010 or E0015	No	Tape Library
Full Document Input File	2	Scratch	No	E0030

Report

Name	Paper Specs	Carriage Tape Specs
Formatted Listing of Document Input	Standard	Standard

Operator Messages

Instructions accompanying halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
01	Date card incorrect	Correct and resubmit. Press Start.
02	Lead card incorrect	Correct and resubmit. Press Start.

03	Date card or lead card missing	Replace both cards. Press Start.
99	End of job	

Operations Data-Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-9	mm/dd/yy

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LEØØ20'
Tape option	7 9	'YES' if output tape is desired, blank if none.
Report option	10-12	'IN ' if input tape is to be printed, 'OUT' if output tape is to be printed, blank if no report desired.
Double spacing option	13-15	'YES' if report is to be double-spaced, blank if single-spaced.
Continuous-form option	16-18	'YES' if report is to be continuous, blank if only one document per page is desired.
Input Index number	19-23	'INPUT INDEX X' where X is the desired number. Not to exceed nine issues per year.

E0030

Function: E0030 will edit the Document Input File for such errors as sequence, coding, format, etc. and list the detected errors.

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program E0030	N/R	Program Library
Lead Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Full Document Input File	1	E0030	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages
Refer to Program Halt List

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions
Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead card not = 'LE0030'	Correct and resubmit. Press Start.
99	End of job	

Operations Data-Lead Card

Field	Cols	Contents
Code	1-6	'LE0030'
Date	7-14	mm/dd/yy

E0040

Function: To update the Document Input File with the contents of the Document Revision File, making additions, deletions, and changes as instructed in the revision input

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program E0040	N/R	Program Library
Lead Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Prior Document Input File	1	Program Library as a result of a prior Editing and Validation cycle E0010 or E0015	Yes	Tape Library
Sorted Document Revisions File	2	Scratch	No	Tape Library
Updated Document Input File	3	Scratch	No	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Document Input Changes	Standard	Standard

Operator Messages
Refer to Program Halt List

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead card not = 'LE0040'	Correct and resubmit. Press Start.
02	Report option not = 'YES' or blank	Correct and resubmit. Press Start.
99	End of job	

Special Instructions

The prior Document Input File was created during the previous Editing and Validation Cycle. It will be updated with the tape from the current Input Processing Cycle.

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LE0040'
Prior Tape	7- 9	'YES' if prior tape, blank if none.
Report Option	10-12	'YES' if report, blank if none.

File Maintenance Module

MAINT1

Function: To format the descriptors, suddescriptors and subject headings for validation against the Thesaurus File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program MAINT1	N/R	Program Library
CFS Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Look-Up Input File	2	Scratch	No	SORTM1
Docu-To-Tape File	3	Scratch	No	MAINT3
Document Input File	4	E0030 or E0040	No	Tape Library

Card Output

File Name	Output Stacker	Disposition
Sequence Errors	N/R (normal punch)	P0010
SORTM1 Control Card	8/2	SORTM1

Report

Name	Paper Specs	Carriage Tape Specs
Listing of Sequence Errors	Standard	Standard

Operator Messages

Sense Switch Settings

Tape Mounting Instructions for the next run

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
E	Tape input

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
001	Date card missing or incorrect	Correct and resubmit. Press Start.
999	End of job	

Special Instructions

The operator messages will mention other sense switches. These options are not used since they reflect other possible input tape formats.

Operations Data CFS Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-7	mmddyy

SORTM1

Function: To sort the descriptors, subdescriptors and subject headings into term sequence

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
SORTM1 Control Card	N/R	Discard
SORT7 Part 2	N/R	Program Libarry

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Look-Up Input File	3	MAINT1	No	Tape Library
Work Tapes	1, 2, 4,	Scratch	No	Scratch
Sorted Look-Up Input File		Scratch	No	MAINT2

(The Sorted Look-Up Input File will be created on either Work Tapes 1 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors on Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Instructions

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-SORTM1 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'31'
	4- 5	'24'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ56'
Input blocking	13-15	'Ø2Ø'
Output blocking	16-18	'Ø2Ø'
Tape reading mode	19-20	'P1'
Header/trailer labels	21-25	'11111'
Padding	26	'9'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'Ø35'
Control field 1	33-36	'ØØ22'
Size of field 1	37-39	'Ø35'
Identification	76-80	'MAIN1'

MAINT2

Function: To validate the descriptive terms against the Thesaurus File

Card Input

File Name	Output Stacker	Disposition
Program MAINT2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Sorted Look-Up Input File	2	SORTM1	No	Tape Library
Thesaurus File	3	Tape Library	Yes	Tape Library
Bad Look-Up File	4	Scratch	No	SORTM2

Card Output

File Name	Output Stacker	Disposition
SORTM2 Control Card	8/2	SORTM2

Operator Messages

Program Halts

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Refer to Operator Messages

SORTM2

Function: To sort the Bad Look-Up File from alphabetical sequence by term into the document and unit number sequence

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
SORTM2 Control Card	N/R	Discard
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Bad Look-Up File	4	MAINT4	No	Tape Library
Work Tapes	1, 2, 3,	Scratch	No	Scratch
Sorted Bad Look-Up File		Scratch	Yes	MAINT3

(The Sorted Bad Look-Up File will be created on either Work Tapes 1 or 3. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors or Wrong-Length Records	NP	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tape

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-SORTM2 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'42'
	4- 5	'13'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ59'
Input blocking	13-15	'ØØ1'
Output blocking	16-18	'ØØ1'
Tape reading mode	19-20	'P1'
Header/trailer labels	21-25	'11111'
Padding	26	'9'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'Ø18'
Control field 1	33-36	'ØØØ1'
Size of field 1	37-39	'Ø18'
Identification	76-80	'MAIN2'

MAINT3

Function: To remove the invalid descriptive terms, substitute the authorized terms for synonyms and geographic terms, and check for additional errors.

Card Input

File Name	Output Stacker	Disposition
Program MAINT3	N/R	Program Library

Tape Set-Up

Field Name	Tape Unit	Source	File Protect	Disposition
Docu-to-Tape File	2	MAINT1	No	Tape Library

Sorted Bad Look-Up File	4	SORTM2	No	Tape Library
Edit File	3	Scratch	No	MAINT4

Card Output

File Name	**Output Stacker**	**Disposition**
Error Cards	NP	P0010

Report

Name	**Paper Specs**	**Carriage Tape Specs**
Error Listing	Standard	Standard

Operator Messages

Sense Switch Settings
Tape Mounting Instructions for next run

Sense Switch Settings

Switch	**Effect of Setting**
A	Last card test
B	To go to end of job

Loading Instructions

Card Load

Program Halt List

Number (B–star)	**Reason**	**Re-start**
002	Tape read error on tape unit 2	Press Start to re-try.
003	Tape write error on tape unit 3	Press Start to re-try.

Special Instructions

Operator messages indicate other sense switch settings. These are not used when terms are validated against the Thesaurus File

Standard Sense Switch Settings: A.

Follow Operator Messages and set Sense Switch B when instructed.

There should be a scratch tape on the fourth tape unit. It should be selected as tape unit 1 but should not be in ready condition.

At the end of the run, the program will come to a halt and will have operator messages on the printer. Follow these messages and press Start. When the program hangs up trying to select tape unit 1, the operator is to end the run and continue on through MAINT4.

MAINT3 – Direct-To-Master File

Function: To format the input to the Document Master File after having bypassed the validation of descriptors, subdescriptors and subject headings

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program MAINT3	N/R	Program Library
CFS Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Input File	2	E0030 or E0040	No	Tape Library
Edit File	3	Scratch	No	MAINT4

Card Output

File Name	Output Stacker	Disposition
Error Cards	N/P	P0010

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages

Sense Switch Settings
Tape Mounting Instructions for next run

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	To go to end of job
E	Tape input

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
001	The date card is missing	Insert date card and press Start.
002	Tape read error on tape unit 2	Press Start to re-try.
003	Tape write error on tape unit 3	Press Start to re-try

Special Instructions

Operator messages indicate other sense switch settings. These are used when input is on cards.

Standard Sense Switch Settings: A, E.

Follow Operator Messages and set Sense Switch B when instructed.

There should be a scratch tape on the fourth tape unit. It should be selected as tape unit 1 but should not be in ready condition.

At the end of run, program will come to a halt and will have operator messages on the printer. Follow these messages and press Start. When the program hangs up trying to select tape unit 1, the operator is to end the run and continue on through MAINT4.

Operations Data-CFS Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-7	mmddyy

MAINT4

Function: Creating and updating the Document Master File

Card Input

File Name	Output Stacker	Disposition
Program MAINT4	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Prior Document Master File	2	Tape Library	Yes	Tape Library
Edit File	3	MAINT3	Yes	Tape Library
Print File	4	Scratch	No	MAINT6
Updated Document Master File	1	Scratch	No	Tape Library

Operator Messages

Operator Instructions

Sense Swtich Settings

Switch	Effect of Setting
A	Last card test
C	Create file first time

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
001	End of load program phase	Dismount monitor, mount scratch tape, press Start.
002	Read error on tape unit 2	Press Start to re-try.
003	Read error on tape unit 3	Press Start to re-try
004	Tape write error on unit drive 4	Press Start to re-try

1111	End of reel on tape unit 1	Mount scratch tape to become next reel of the Document Master File, press Start.
1234	Delete failure	Disaster error. Restart run.
2121	Tape label error on tape unit 2	Incorrect file mounted on unit drive 2. Mount the correct Document Master File, press Start.
3131	Tape label error on tape unit 3	Mount the Edit File, press Start.
4141	End of reel on tape unit 4	Mount scratch tape to become next reel of the Print File, press Start.
4444	Bad tape on tape unit 1 for the updated Document Master File	Restart the run with a tape that has a tape mark on it.

Special Instructions

Standard Sense Switch Settings: A

Do not use a new tape for the updated Document Master File.

Tape unit 1 is unloaded at beginning of program. Ignore operator message and ready unit 1

At end of run, program will come to a halt and will have operator messages on printer.

MAINT5

Function: To provide tape labelling instructions for output of MAINT4 and set-up instructions for MAINT6

Card Input

File Name	Output Stacker	Disposition
Program MAINT5	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Print File	4	MAINT4	No	Tape Library
Work Tape	2, 3	Scratch	No	MAINT6

Operator Messages

Tape labelling instructions for output of MAINT4

Set-up instructions for MAINT6

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

MAINT6

Function: To provide an audit trail of the maintenance cycle and to create the means of updating the Inverted File

Card Input

File Name	Output Stacker	Disposition
Program MAINT6	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Print File	4	MAINT4	No	Tape Library
Work Tape	2	Scratch	No	Scratch
Descriptor Input File	3	Scratch	No	SORTD1

Card Output

File Name	Output Stacker	Disposition
SORTD1 Control Card	8/2	SORTD1

Report

Name	Paper Specs	Carriage Tape Specs
Transactions to the Document Master File, Non-Processed Transactions	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	Print Docu/Adds Deletes
C	Print Docu/Adds Changes
D	Print Docu/Error Changes
F	Bypass Master File Activities
G	Proceed to end of job

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
004	Tape write error on tape unit 3	Press Start to re-try
222	Label error on tape unit 2	Mount a new scratch on tape unit 2, press Start.
333	Label error on tape unit 3	Mount a new scratch on tape unit 3, press Start.
444	Label error on tape unit 4	Mount a new scratch on tape unit 4, press Start.
3131	End of reel on tape unit 3	Mount a scratch tape to become the next reel of the Descriptor Input

4141	End or reel on tape unit 4	File, press Start. Mount next reel of the Print File, press Start.

Special Instructions

Standard Sense Switch Settings:
Sense Switch A;
Sense Switches A and D. Press Start;
Sense Switches A and G. Press Start.

For restart in case of run failure:
Sense Switch A;
Sense Switches A and F. Press Start;
Sense Switch A. Press Start;
Sense Switches A and D. Press Start;
Sense Switches A and G. Press Strart.

MAINT7

Function: To generate new Descriptor Input File and/or merge two Document Master Files

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program MAINT7	N/R	Program Library
CFS Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
New Document Master File	4	Tape Library	No	Tape Library
Prior Document Master File	2	Tape Library	Yes	Tape Library
Descriptor Input File	3	Scratch	No	SORTD1
Merged Document Master File	1	Scratch	No	Tape Library

Card Output

File Name	Output Stacker	Disposition
SORTD1 Control Card	N/P	SORTD1

Operator Messages

Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
C	One Document Master File
D	No descriptor output

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
022	Prior Document Master File out of sequence	Sort the file and restart.
044	New Document Master File out of sequence	Sort the file and restart.

Special Instructions

Sense Switches A and C to:
Duplicate Document Master File
Re-generate Descriptor Input File.

Sense Switches A and D to:
Merge two Document Master Files.

Operations Data-CFS Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-7	mmddyy

MAINT8

Function: To print the Statistical Analysis of the Document Master File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program MAINT8	N/R	Program Library
Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Master File	2	Tape Library	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Unauthorized Segments Statistical Analysis	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (I-star)	Reason	Re-start
673	Tape write error	Press Start to try ten times more.

695	Tape read error	Press Start to try ten times more.
890	Date card missing or incorrect	Correct and resubmit. Press Start.
949	File out of sequence	Sort file and restart.
1423	End of job	

Operations Data-Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-9	mm/dd/yy

MAINT9

Function: To generate a Document Master Subset File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program MAINT9	N/R	Program Library
Finder Cards	1	Cards out of sequence. Rerun. Program Library

Tape Set Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Master File	2	Tape Library	Yes	Tape Library
Document Master Subset File	3	Scratch		Tape Library

Operator Messages

Error Messages

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Resaon	Re-start
222	Tape read error	Press Start to try ten times more.
333	Tape write error	Press Start to try ten times more.
999	End of job	

Special Instructions

If message appears indicating errors, correct them and rerun the job.

Operations Data-Finder Cards

Field	Cols	Contents
Code	1	'F'
Significant portions of selected document numbers	2-15	The left-most characters of a document number, as many as are significant.

LIBPRT

Function: To print a listing of the Document Master File

Card Input

File Name	Output Stacker	Disposition
Call Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
CFS Systems Tape	1	Tape Library	Yes	Tape Library
Document Master File or output of MAINT9	4	Tape Library or MAINT9	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Library Print	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Tape Load, Start

Program Halt List

Number (B-star)	Reason	Re-start
004	Tape write error on tape unit 3	Press Start to re-try.
222	Label error on tape unit 2	Mount a new scratch on tape unit 2, press Start.
333	Label error on tape unit 3	Mount a new scratch on tape unit 4, press Start.
444	Label error on tape unit 4	Mount a new scratch on tape unit 4, press Start.
3131	End of reel on tape unit 3	Mount a scratch tape to become the next reel of the Descriptor Input File, press Start.
4141	End of reel on tape drive 4	Mount next reel of the Document

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program Name	8-13	'LIBPRT'

SORTD1

Function: To sort the Descriptor Input File into alphabetical sequence by term to update the Inverted File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
SORTD1 Control Card	N/R	Discard
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Descriptor Input File	3	MAINT6	No	Tape Library
Work Tapes	1, 2, 3	Scratch	No	Scratch
Sorted Descriptor Input File			No	DMAIN1

(The Sorted Descriptor Input File will be created on either Work Tapes 1 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/R	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Insrructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-SORTD1 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'31'
	4- 5	'24'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ42'
Input blocking	13-15	'ØØ4'
Output blocking	16-18	'ØØ4'
Tape reading mode	19-20	'P1'
Header/trailer labels	21-25	'11111'
Padding	26	'9'
Core storage size	27	'4'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'Ø39'
Control field 1	33-36	'ØØØ4'
Size of field 1	37-39	'Ø39'
Identification	75-80	'MAIN6'

DMAIN1

Function: To create/update the Inverted File

Card Input

File Name	Output Stacker	Disposition
Program DMAIN1	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Sorted Descriptor Input File	2	SORTD1	No	Tape Library
Prior Inverted File	3	Tape Library	Yes	Tape Library
Updated Inverted File	4	Scratch	No	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Errors	Standard	Standard

Operator Messages

Sense Switch Settings
Operator Instructions

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
C	Create file first time

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
2121	End of reel on tape unit 2	Mount the next reel of the Sorted Descriptor Input File, press Start.
3131	End of reel on tape unit 3	Mount the next reel of the Inverted File, press Start.
222	Wrong tape mounted on unit 2	Mount the Sorted Descriptor Input File on unit 2, press Start.
333	Wrong tape mounted on unit 3	Mount the Prior Inverted File on unit 3, press Start.
1953	Sorted Descriptor Input File out of sequence	Perform SORTD1 and restart the run.

Special Instructions

Standard Sense Switch Settings: A.

When all tapes are unloaded, the operator is to consider this as a termination of the run.

DMAIN2

Function: To print the Inverted File Summary Listing by category: the entire file; the subject analysis terms; or the internal descriptors and document numbers.

Card Input

File Name	Output Stacker	Disposition
Program DMAIN2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Inverted File	4	DMAIN1	Yes	Tape Library
Work Tape	2	Scratch	No	Scratch

Report

Name	Paper Specs	Carriage Tape Specs
Summary Listing	Standard	Standard

Operator Messages

Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
C	To produce the subject analysis terms
D	To produce the document number listing

Loading Instructions
Card Load

Program Halt List

Number (B-star)	Reason	Re-start
002	End of reel on tape drive 4	Mount the next reel of the Inverted File, press Start.
611	Tape read error on work rape	Press Start to re-try.
612	Tape write error on work tape	Press Start to re-try.

Special Instructions
Standard Sense Switch Settings:
Sense Switches A and C;
Sense Switches A and D.

If the document number listing is desired, the job will be run a second time with Sense Swtiches A and D.

DMAIN3

Function: To print the entire Inverted File or only the subject analysis terms

Card Input

File Name	Output Stacker	Disposition
Program Deck DMAIN3	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Inverted File	4	DMAIN1	Yes	Tape Library
Work Tape	2	Scratch	No	Scratch

Report

Name	Paper Specs	Carriage Tape Specs
Detail Listing	Standard	Standard

Operator Messages
Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
C	To produce expanded report
D	To produce the subset of subject terms

Loading Instructions
Card Load

Program Halt List

Number (B-star)	Reason	Re-start
011	Tape error writing work tape	Press Start to re-try
012	Tape error reading work tape	Press Start to re-try
002	End of reel on tape drive 2	Mount the next reel of the Inverted File, press Start.

Special Instructions
Standard Sense Switch Settings: A, D.

Search Module

SEP1

Function: To expand the Boolean statement of a search request

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program SEP1	N/R	Program Library
Search Input	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Table File	2	Scratch	No	SORT SEP1
Request File	3	Scratch	No	SEP3

Card Output

File Name	Output Stacker	Disposition
SORT SEP1 Control Card	N/P	SORT SEP1

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing and Messages	Standard	Standard
Optional Printout and Expanded Form Listing	Standard	Standard

Operator Messages
Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	Printout the Expanded Form

Loading Instructions
Card Load

Program Halt List

Number (B-star)	Reason	Re-start
222	Write Tape Error (TU2)	Press Start to try ten times more, else restart with new scratch.
333	Write Tape Error (TU3)	Press Start to try ten times more, else restart with new scratch.
999	End of job	

SORT SEP1

Function: To sort the descriptors used in the request into alphabetical sequence

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
SORT SEP1 Control Card	N/R	Discard
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Table File	2	SEP1	Yes	Tape Library
Work Tapes	1, 3, 4	Scratch	No	Scratch
Sorted Table File		Scratch	No	SEP2

(The Sorted Table File will be created on either Work Tapes 1 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Switch
A	Last card test
D	Save the input tape

Loading Instructions
Card Load

Program Halt List
See SORT7 Halt List at the end of this chapter

Operations Data-SORT SEP1 Control Card 1

Field	Cols	Contents
Tape drive assignments	1- 2	'21'
	4- 5	'34'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ9Ø'
Input blocking	13-15	'ØØ1'
Output blocking	16-18	'ØØ1'
Tape reading mode	19-20	'P2'
Core storage size	27	'5'
Number of control fields	28-29	'Ø2'
Number of control characters	30-32	'Ø32'
Control field 1	33-36	'ØØØ4'
Size of field 1	37-39	'ØØ1'
Identification	76-80	'SSEP1'

Operations Data-SORT SEP1 Control Card 2

Field	Cols	Contents
Control field 2 location	1- 4	'ØØØ6'
Control field 2 length	5- 7	'Ø36'
Identification	76-80	'SSEP1'

SEP2

Function: To append frequencies of usage to terms

Card Input

File Name	Output Stacker	Disposition
Program SEP2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Inverted File	2	Tape Library	Yes	Tape Library
Sorted Table File	3	SORT SEP1	No	Tape Library
Table File w/Frequencies	4	Scratch	Yes	SORT SEP2

Card Output

File Name	Output Stacker	Disposition
SORT SEP2 Control Card	N/P	SORT SEP2

Report

Name	Paper Specs	Carriage Tape Specs
Error List and Messages	Standard	Standard

Operator Messages

Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	To adapt system for URBANDOC's use

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
222	Read tape error on (TU2)	Press Start to read again.
333	Read tape error on (TU3)	Press Start to read again.
444	Write tape error on (TU4)	Press Start to write again, else restart with new scratch.
999	End of job	

SORT SEP2

Function: To sort the Table File with Frequencies

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
SORT SEP2 Control Card	N/R	Discard
SORT 7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Table File w/Frequencies	4	SEP2	Yes	Tape Library
Work Tapes	1, 2, 3	Scratch	No	Tape Library
Sorted Table File w/ Frequencies		Scratch	No	SEP3

(The Sorted Table File w/Frequencies will be created on either Work Tapes 1 or 3. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/R	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard 1	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last Card test
D	Save the input tape

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of the chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data—SORT SEPT2 Control Card 1

Field	Cols	Contents
Tape drive assignments	1- 2	'41'
	4- 5	'23'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ9Ø'
Input blocking	13-15	'ØØ1'
Output blocking	16-18	'ØØ1'
Tape reading mode	19-20	'P2'
Core storage size	27	'5'
Number of control fields	28-29	'Ø3'
Number of control characters	30-32	'Ø18'
Control field 1	33-36	'ØØ81'
Size of field 1	37-39	'ØØ9'
Identification	76-80	'SSEP2'

Operations Data—SORT SEP2 Control Card 2

Field	Cols	Contents
Control field 2 location	1- 4	'ØØØ1'
Control field 2 length	5- 7	'ØØ4'
Control field 3 location	8-11	'ØØ71'
Control field 3 length	12-14	'ØØ5'
Identification	76-80	'SSEP2'

SEP3

Function: Final expansion of Boolean statement:
Requests punched out as written on tape; and
Printout of requests

Card Input

File Name	Output Stacker	Disposition
Program SEP3	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Sorted Table File w/ Frequencies	2	SORT SEP2	Yes	PHASE6, PASS1

Operating Instructions

Request File	3	SORT SEP1	Yes	Tape Library
CFS Request Input File	1	Scratch	No	PHASE1

Card Output

File Name	Output Stacker	Disposition
CFS Requests	N/P	PHASE1

Report

Name	Paper Specs	Carriage Tape Specs
Listing of error mesages	Standard	Standard

Operator Messages

Sense Switch Settings

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
B	On: tape output Off: card output

Program Halt List

Number (B-star)	Reason	Re-start
111	Tape write error (TU1)	Press Start to try rewriting.
222	Read tape error (TU2)	Press Start to try re-read.
333	Read tape error (TU3)	Press Start to try re-read.
999	End of job	

Special Instructions

Standard Sense Switch Settings: A.

SEARCH PHASE1

Function: To format and edit the requests

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Call Card	N/R	Program Library
CFS Requests	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Systems Tape, CFS Requests	1	Tape Library	Yes	PHASE2
Scratch	2	Scratch	No	PHASE2
CFS Request File	3	Scratch	No	PHASE2
Scratch	4	Scratch	No	PHASE2

Operator Messages

Error Messages
Tape Mounting Instructions for the next run
Restart and Continuation Procedures

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Tape Load, Start

Program Halt List

Number (B-star)	Reason	Re-start
050	Tape read error on systems tape	Press Start to re-try.
004	Tape write error on drive 4	Press Start to re-try.
003	Tape write error on drive 3	Press Start to re-try.
002	Tape write error on drive 2	Press Start to re-try.
5431	Tape read error on drive 2	Press Start to re-try. Else re-start PHASE1 with a new scratch tape.
5872	Tape write error on drive 3	Press Start to re-try. Else re-start PHASE1 with a new scratch tape.

Special Instructions

If tape input, after program load a halt will occur and the CFS System Tape is dismounted. It is replaced by the CFS Request File. Press Start to continue processing.

After editing the search requests, there is a halt. At this point the user may either continue the search or end the run.

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PHASE1

SEARCH PHASE2

Function: Primary Document Selection

Card Input

If processing is continuous, no card input
If restart after interrupt, use Call Card

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Systems Tape	1	Tape Library	Yes	PHASE3
Inverted File	2	Tape Library	Yes	Tape Library
CFS Request File	3	PHASE1	Yes	PHASE4
Primary Selection File	4	Scratch	No	PHASE3

Operator Messages

Halt Messages
Tape Mounting and Labelling Instructions
Restart and Continuation Procedures

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

If processing is not interrupted after the completion of PHASE 1, PHASE 1 will internally load the PHASE 2 program. Press Start.

If processing is interrupted after the completion of PHASE 1, press Tape Load and Start

Program Halt List

Number (B-star)	Reason	Re-start
002	Tape read error on tape drive 2	Press Start to re-try. Else must restart with a new Inverted File.
004	Tape write error on tape drive 4	Press Start to re-try or restart with a new scratch.
050	Tape read error for CFS Systems Tape	Press Start to re-try.

Special Instructions

Under standard conditions processing will be continuous.

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PHASE2'

SEARCH PHASE3

Function: To sort document list of the Primary Selection File into document number sequence

Card Input

If processing is continuous, no card input
If restart after an interrupt, use Call Card

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Systems Tape	1	Tape Library	Yes	PHASE4
Reduced Primary Selection File	2	Scratch	No	PHASE4
Work Tape	3	Scratch	No	Scratch
Primary Selection File	4	PHASE2	Yes	Tape Library

Operator Messages

Halt Messages
Tape Mounting Instructions for the next run
Restart and Continuation Procedures

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

If processing is not interrupted after the completion of PHASE2, PHASE2 will internally load the PHASE3 program. Press Start.

If processing is interrupted after the completion of PHASE2, press Tape Load and Start.

Program Halt List

Number (B-star)	Reason	Re-start
050	Tape read error on CFS Systems Tape	Press Start to re-try.
444	Wrong tape on drive 4	Mount the Primary Selection File on drive 4, press Start.
2233	Tape read error on drive 2 or drive 3	Press Start to re-try.

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PHASE3'

SEARCH PHASE4

Function: To associate the request terms with the primary selected document numbers

Card Input

If processing is continuous, no card input
If restart after an interrupt, use Call Card

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
System Tape	1	Tape Library	Yes	PHASE5
Reduced Primary Selection File	2	PHASE3	Yes	Tape Library
Request File	3	PHASE1	Yes	PHASE6
Inquiry File	4	Scratch	No	PHASE5

Disk Input

File Name	Disk Unit	Disposition
Scratch	0	Scratch

Operator Messages

PHASE4 Options
Halt Messages
Tape Mounting Instructions for the next run
Restart and Continuation Procedures

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
G	Use the 1311 disk drive with a scratch pack for this program

Loading Instructions

If processing is not interrupted after the completion of PHASE3, PHASE3 will internally load the PHASE4 program. Press Start

If processing is interrupted after the completion of PHASE3, press Tape Load and Start

Program Halt List

Number (B-star)	Reason	Re-start
050	Tape read error on CFS Systems Tape, drive 1	Press Start to re-try.
004	Tape write error on drive 4	Press Start to re-try.
900	Tape write error	Press Start to re-try.
901	Tape read error	Press Start to re-try.

Special Instructions

Under standard conditions, processing will be continuous

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PHASE4'

SEARCH PHASE5

Function: Secondary document selection

Card Input

If processing is continuous, no card input
If restart after an interrupt, use Call Card

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
System Tape	1	Tape Library	Yes	PHASE6
Selection File	2	Scratch	No	PHASE6
Document Master File	3	Tape Library	Yes	Tape Library
Inquiry File	4	PHASE4	No	Tape Library

Operator Messages

Halt Messages
Tape Mounting Instructions for the next run
Restart and Continuation Procedures

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

If processing is not interrupted after the completion of PHASE4, PHASE4 will internally load the PHASE5 program. Press Start

If processing is interrupted after the completion of PHASE4, press Tape Load and Start

Program Halt List

Number (B-star)	Reason	Re-start
050	Tape read error on CFS Systems Tape, drive 1	Press Start to re-try.
333	Wrong tape on drive 3	Mount the Document Master File on drive 3, press Start.
444	Wrong tape on drive 4	Mount the Inquiry File on drive 4, press Start.
2121	End of reel on Selection File	Press Start to continue with only partial output. For complete output, restart the run with a larger tape.
2233	Tape write error on drive 2 or drive 3	Press Start to re-try.

Special Instructions

Under standard conditions, processing will be continuous

Operations Data — Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PHASE5'

SEARCH PHASE6, PASS1 (BASIC)

Function: Print Output from Search Run

Card Input

If processing is continuous, no card input
If restart after an interrupt, use Call Card

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Systems Tape	1	PHASE5	Yes	Tape Library
Selection File	2	PHASE5	No	PHASE6, PASS2
CFS Request File	3	PHASE1	Yes	PHASE6, PASS2
Work Tape	4	Scratch	No	PHASE6, PASS2

Operator Messages

Halt Messages
End of PASS1 message
Tape Mounting Instructions for PASS2

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

If processing is not interrupted after the completion of PHASE5, PHASE5 will internally load the PHASE6 program. Press Start

If processing is interrupted after the completion of PHASE5, press Tape Load and Start

Program Halt List

Number (B-star)	Reason	Re-start
004	Tape write error on drive 4	Press Start to re-try.
050	Tape read error on CFS Systems Tape, drive 1	Press Start to re-try.

Operations Data-Call Card

Field	Cols	Contents
Code	1- 4	'CALL'
Program name	8-13	'PH6PA1'

SEARCH PHASE6, PASS2 (BASIC)

Function: Print Output from Search Run

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Scratch	1	Scratch	No	Scratch
Selection File	2	PHASE5	No	Tape Library
CFS Request File	3	PHASE1	Yes	Tape Library
Work Tape	4	PHASE6, PASS1	No	Scratch

Report

Name	Paper Specs	Carriage Tape Specs
Retrieval Report	3-ply paper	*Retrieval Report* Tape 1 or Tape 2

Operator Messages

Halt Messages
End of Run Message

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Continuous from PHASE6, PASS1

Program Halt List

Number (B-star)	Reason	Re-start
222	Wrong tape on drive 2	Mount Selection File on drive 2, press Start.

Special Instructions

Retrieval Report Tape 1 is for a continuous bibliography. Prepare the carriage tape with the following specifications:

Channel 1: Heading of the page;
Channel 12: Last line of the page.

Retrieval Report Tape 2 is for a bibliography with one reference per page. Prepare the carriage tape with the following specifications:

Channel 1: Heading of the page;
Channel 9: Three lines below Channel 1;
Channel 12: Last line of the page.

SEARCH PHASE6, PASS1 (EXPANDED)

Function: To print output from Search Run

Card Input

File Name	Output Stacker	Disposition
Program PHASE6, PASS1	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Sorted Table File w/Frequencies	1	SORT SEP2	Yes	Tape Library
Selection File	2	PHASE5	Yes	PHASE6, PASS2
Request File	3	PHASE1	Yes	PHASE6, PASS2
Work Tape	4	Scratch	No	PHASE6, PASS2

Operator Messages

Halt Messages
End of PASS1 message
Tape Mounting Instructions for PASS2

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List

Number (B-star)	Reason	Re-start
004	Write error on tape drive 4	Press Start to try again

Special Instructions
Go directly to PASS2. *DO NOT RESET TAPE DRIVES*
End of run indicated by printer message
Must replace System Tape with Sorted Table File w/Frequencies

SEARCH PHASE6, PASS2 (EXPANDED)

Function: To print the output from Search Run

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program PHASE6, PASS2	N/R	Program Library
CFS Date Card	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Scratch	1	Scratch	No	Scratch
Selection File	2	PHASE5	Yes	Tape Library
Request File	3	PHASE1	Yes	Tape Library
Work Tape	4	PHASE6, PASS 1	No	Scratch

Card Output

File Name	Output Stacker	Disposition
Search Summary Report	N/P	Off-line storage

Report
Retrieval Report

Operator Messages
Halt Messages
End of job message

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions
Card Load

Program Halt List

Number (B-star)	Reason	Re-start
222	Wrong tape on drive 2	Mount correct tape, press Start.
333	Tape write error	Press Start to try to write ten times more.
999	End of job	

Special Instructions
Do not reset tapes after PASS1

Operations Data-CFS Date Card

Field	Cols	Contents
Code	1	'D'
Date	2-7	mmddyy

Publication Module

P0010

Function: To create the Publication File and an optional Content Analysis File

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0010	N/R	Program Library
Lead Card	N/R	Discard
Error Cards	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Document Input File	1	Tape Library	Yes	Tape Library
Publication File	2	Scratch	No	P0025
Content Analysis File	3	Scratch	No	P0015

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages

Halt Message

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead Card missing or wrong	Replace card, press Start.
99	End of job	

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LP0010'
Element numbers	7-46	A maximum of twenty element numbers, two positions each, i.e., '010299'

P0015

Function: To sort the Content Analysis File into sequence by descriptor within document number

Card Input

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
P0015 Control Card	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Content Analysis File	3	P0010	No	Tape Library
Work Tapes	1, 2, 4	Scratch	No	Scratch
Sorted Content Analysis File		Scratch		P0050

(The Sorted Content Analysis File will be created on either Work Tapes 1 or 4. There will be an operator message at the end of the job indicating which tape unit contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Halt List

See the SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-P0015 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'31'
	4- 5	'24'

Number of input reels	7- 8	'01'
Record length	9-12	'0038'
Input blocking	13-15	'010'
Output blocking	16-18	'010'
Tape reading mode	19-20	'P1'
Core storage size	27	'5'
Number of control fields	28-29	'01'
Number of control characters	30-32	'037'
Control field 1	33-36	'0001'
Size of field 1	37-39	'037'
Identification	76-80	'P0015'

P0025

Function: To sort the Publication File into document number, unit number, element number sequence

Card Input

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
P0025 Control Card	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Publication File	1	P0010	No	Scratch
Work Tapes	2, 3, 4	Scratch	No	Scratch
Sorted Publication File		Scratch	No	P0040 or P0050

(The Sorted Publication File will be created on either Work Tapes 2 or 4. There will be an operator message at the end of the job indicating which tape unit contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-P0025 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'12'
	4- 5	'34'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'Ø4Ø9'
Input blocking	13-15	'ØØ3'
Output blocking	16-18	'ØØ3'
Tape reading mode	19-20	'P1'
Padding	26	'9'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'117'
Control field 1	33-36	'ØØ59'
Size of field 1	37-39	'117'
Record mark padding	74	'1'
Identification	76-80	'P0025'

P0040

Function: To create a subset of the Publication File in which there will be a citation of main entry and title for each subject heading assigned to a document

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0040	N/R	Program Library
Lead Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Publication File	1	Tape Library	Yes	Tape Library
Work Tape	2	Scratch	No	Scratch
Subject Subset File	4	Scratch	No	P0045

Report

Name	Paper Specs	Carriage Tape Specs
Error Listing	Standard	Standard

Operator Messages

Refer to Program Halt List

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead card not = 'LP0040'	Correct and re-enter card. Press Start
02	Change address of tape drive 2 to tape drive 3	Press Start after unit address is changed.
99	End of job	

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LP0040'
Element number	7-36	A maximum of ten element numbers, three positions each, i.e., '001011'

P0045

Function: This sort will sequence the Subject Subset File by subject heading, document number and element number

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
P0045 Control Card	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Subject Subset File	1	P0040	Yes	Tape Library
Work Tapes	2, 3, 4	Scratch	No	Scratch
Sorted Subject Subset File		Scratch	No	P0050

(The Sorted Subject Subset File will be created on either Work Tapes 2 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D.
Press Start

Operations Data-P0045 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'12'
	4- 5	'34'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'Ø4Ø9'
Input blocking	13-15	'ØØ3'
Output blocking	16-18	'ØØ3'
Tape reading mode	19-20	'P1'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'
Number of control characters	30-32	'Ø75'
Control field 1	33-36	'ØØØ1'
Size of field 1	37-39	'Ø75'
Identification	76-80	'P0045'

P0050

Function: To print either the Main Document Listing or the Major Subject Listing

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0050	N/R	Program Library
Parameter Card	N/R	Discard
Option Cards	N/R	Discard
Selected Document Cards	N/R	Discard

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Publication File		P0025		
or	1		Yes	Tape Library
Sorted Subject Subset File		P0040		
Content Analysis File	2	P0015	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Main Document Listing	Standard	P0050 Carriage Tape
or		
Major Subject Listing	Standard	P0050 Carriage Tape

Operator Messages

Halt Messages

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding IOCS Halts)

Number (B-star)	Reason	Re-start
01	ID field wrong on option card	Correct and resubmit. Press Start.
02	End of Selected Document Cards	Resubmit. Press Start.
05	Index card missing	Insert card. Press Start.
99	End of job	

Special Instructions

Prepare the P0050 Carriage Tape with the following specifications:

Channel 1: Heading of the page;
Channel 9: Last line of the page;
Channel 12: Page number

Operations Data-Parameter Card

Field	Cols	Contents
Code	1-6	'XP0050'
Title option	7-9	'YES' if title card present, blank if no title.
Page option	10-12	'YES' if lead card present, blank if all elements to be printed.
Tape option	16-18	'YES' if tape output desired, blank if not.
Index type	19-21	'YES' for index card, must be present.
Optional lead	22-24	'YES' if optional lead card present, blank if not.
Selected documents	25-27	'YES' if selected documents cards present, blank if none.

Opertations Data-Title Card

Field	Cols	Contents
Code	1-6	'TP0050'
Title	7-80	

Operations Data-Page Card

Field	Cols	Contents
Code	1-6	'PP0050'
Page number	7-9	Desired page number, minus one, from where printing is to start.

Operations Data-Lead Card

Field	Cols	Contents
Code	1-6	'LP0050'
Element numbers	7-57	A maximum of seventeen element numbers, three positions each, i.e., '001002009'

Operations Data-Tape Option Card

Field	Cols	Contents
Code	1-6	'CP0050' card is present only if the Content Analysis File is used.

Operations Data-Index Card

Field	Cols	Contents
Code	1-6	'IP0050'
Index Type	7-9	'SUB' if Major Subject listing, 'DOC' if Main Document Listing. This card has to be present.

Operations Data-Optional Lead Card

Field	Cols	Contents
Code	1- 6	'OP0050'
Element numbers	7-36	A maximum of ten element numbers each three positions, i.e., '001002099'

Operations Data-Selected Document Card

Field	Cols	Contents
Code	1- 6	'DP0050'
Element number	7-20	Number of desired document to be processed.

P0055

Function: To sequence the Publication File by sort control field and document number

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
P0055 Control Cards	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Publication File	1	Tape Library	Yes	Tape Library
Work Tapes	2, 3, 4	Scratch	No	Scratch
Publication Index File		Scratch	No	P0060

(The Publication Index File will be created on either Work Tapes 2 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tapes

Loading Instructions

Card Load

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start

Operations Data-P0055 Control Card 1

Field	Cols	Contents
Tape drive assignments	1- 2	'12'
	4- 5	'34'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'Ø4Ø9'
Input blocking	13-15	'ØØ3'
Output blocking	16-18	'ØØ3'
Tape reading mode	19-20	'P1'
Core storage size	27	'5'
Number of control fields	28-29	'Ø2'
Number of control characters	30-32	'114'
Control field 1	33-36	'ØØ86'
Size of field 1	37-39	'1ØØ'
Identification	76-80	'PØØ55'

Operations Data-P0055 Control Card 2

Field	Cols	Contents
Control field 2 location	1- 4	'0059'
Control field 2 length	5- 7	'04'
Identification	76-80	'0055'

P0060

Function: To produce the various sections of the *Input Index*, either as a report, a formatted tape or both.

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0060	N/R	Program Library
Option Card	N/R	Program Library
Title Card	N/R	Program Library
Lead Card	N/R	Program Library
Page Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Publication Index File	1	Tape Library	Yes	Tape Library
Formatted Page File	2	Scratch	No	P0065

Report

Name	Paper Specs	Carriage Tape Specs
Sections of the *Input Index*	Standard	P0050

Operator messages

Refer to Program Halt List

Sense Swtich Settings

Swtich	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Option card not = 'OP0060'	Correct and re-enter card. Press Start.
02	Title card not = 'TP0060'	Correct and re-enter card. Press Start
03	Lead card not = 'LP0060'	Correct and re-enter card. Press Start.
04	Page card not = 'PP0060'	Correct and re-enter card. Press Start.

05	Sort control field out of sequence	Sort. Restart job.
99	End of job	

Special Instructions

Refer to P0050 for the preparation of the P0050 carriage tape

Operations Data-Option Card

Field	Cols	Contents
Code	1- 6	'OP0060'
Subject	7- 9	'YES' if there is a Lead Card, blank if none.
Duplicate entries	10-12	'YES' if duplicate lines to be eliminated, blank is not.
Spacing	13-15	'YES' if an extra space is desired between documents.
Indent	16-18	'YES' if document is desired below the text line, blank if not.
Tape	19-21	'YES' if tape is desired, blank if not.
Page	22-24	'YES' if there is a page card, blank if none.
Title	25-27	'YES' if there is a title card, blank if none.
Columns	28-30	'001' if single column page, '002' if double page.

Operations Data-Title Card

Field	Cols	Contents
Code	1- 6	'TP0060
Title	7-80	Any title

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LP0060'
Element numbers	7-51	A maximum of fifteen element numbers, three positions each, i.e., '001015099'

Operations Data-Page Card

Field	Cols	Contents
Code	1- 6	'PP0060'
Index number	7- 9	Index number
Page number	10-12	Page number, less one, where printing to start
Page depth	13-15	Number of lines per page

P0065

Function: To arrange the Formatted Page File by index, page, line and column number

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
SORT7 Part 1	N/R	Program Library
P0065 Control Card	N/R	Program Library
SORT7 Part 2	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Formatted Page File	1	P0065	Yes	Tape Library
Work Tapes	2, 3, 4	Scratch	No	Scratch
Sorted Formatted Page File		Scratch	No	P0070

(The Sorted Formatted Page File will be created on either Work Tapes 2 or 4. There will be an operator message at the end of the job indicating which drive contains the output of the sort.)

Card Output

File Name	Output Stacker	Disposition
Tape Read Errors and Wrong-Length Records	N/P	Discard

Report

Name	Paper Specs	Carriage Tape Specs
SORT7 Operator Messages	Standard	Standard

Sense Switch Settings

Switch	Effect of Setting
A	Last card test
D	Save the input tape

Program Halt List

See SORT7 Halt List at the end of this chapter

Special Instructions

If Sense Switch D is on, there will be a program halt and a message instructing the operator to remove the input and mount a scratch in its place. Turn off Sense Switch D. Press Start.

Operations Data-P0065 Control Card

Field	Cols	Contents
Tape drive assignments	1- 2	'12'
	4- 5	'34'
Number of input reels	7- 8	'Ø1'
Record length	9-12	'ØØ69'
Input blocking	13-15	'Ø1Ø'
Output blocking	16-18	'Ø1Ø'
Tape reading mode	19-20	'P1'
Core storage size	27	'5'
Number of control fields	28-29	'Ø1'

Number of control characters	30-32	'007'
Control field 1	33-36	'0004'
Size of field 1	37-39	'007'
Identification	76-80	'P0065'

P0070

Function: To print the double column sections of the *Input Index*

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0070	N/R	Program Library
Title Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Formatted Page File	1	P0065	Yes	Tape Library

Report

Name	Paper Specs	Carriage Tape Specs
Index Sections	Standard	P0050

Operator Messages

Refer to Program Halt List

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Lead Card not = 'LP0070'	Correct and resubmit. Press Start.
02	Index numbers do not match	Correct and resubmit. Press Start.
03	Page number out of sequence	Correct and resubmit. Press Start.
04	Line number of out sequence	Correct and resubmit. Press Start.
99	End of job	

Operations Data-Title Card

Field	Cols	Contents
Code	1- 6	'LP0070'
Index number	7- 9	
Page number	10-12	

P0080

Function: To create a Special Publication File from either the Selection File of the search run or from the Document Master File.

Card Input (in deck set-up order)

File Name	Output Stacker	Disposition
Program P0080	N/R	Program Library
Lead Card	N/R	Program Library

Tape Set-Up

File Name	Tape Unit	Source	File Protect	Disposition
Selection File or Document Master File	1	PHASE5	Yes	Tape Library
Special Publication File	2	Tape Library Scratch	No	P0025

Operator Messages

Halt Messages

Sense Switch Settings

Switch	Effect of Setting
A	Last card test

Loading Instructions

Card Load

Program Halt List (excluding the IOCS Halts)

Number (B-star)	Reason	Re-start
01	Wrong Lead Card	Correct and resubmit. Press Start.
02	Selection File not mounted	Mount Selection File. Press Start.
03	Document Master File not mounted	Mount Document Master File. Press Start.
05	Lead Card missing	Insert Lead Card. Press Start.
99	End of job	

Operations Data-Lead Card

Field	Cols	Contents
Code	1- 6	'LP0080'
Tape type	7- 9	'SEL' if Selection File is used as input, 'DOC' if Document Master File is used as input.
Element numbers	10-69	A maximum of twenty element numbers, each three positions, i.e., '001002099'

SORT7 Halt List[1]

Number	I-Address Register	Sequence Number	Message and/or Reason	Procedure
1.	1034	1371	RECORD LENGTH ERROR - MAXIMUM LESS THAN MINIMUM - RESTART Reason: The specified maximum record length is less than the specified minimum record length.	Correct the control card and restart the sort.
2.	1226	1402	INVALID LABEL SPECIFICATION Reason: The user has specified a combination of 80-character and 120-character labels in columns 21-25 of control card 1.	Either press the start key for 80-character labels, or press the start reset and start keys for 120-character labels.
3.	2542	591	CF LNGH ERROR - FOR COMPUTED TLCF, PRESS START Reason: The total control-data field length specified in columns 30-32 of control card 1 is not equal to the sum of the lengths of the individual control-data fields.	Either press the start key to accept the total control-data field length computed by the program or correct the control card and restart the sort.
4.	3636	771	MACHINE SIZE ERROR - FOR 8K, PRESS START Reason: The system core-storage capacity specified in column 27 of control card 1 is incorrect.	Either press the start key if the system capacity is 8,000 positions of core storage, or correct the control card and restart the sort.
5.	4210	835	OVERSIZE INPUT BLOCKING Reason: The input block length for variable length specified in columns 13-15 of control card 1 is greater than the maximum allowable block length.	Correct the control card and/or condition and restart the sort.
6.	4229	841	OVERSIZE RECORD LENGTH - RESTART Reason: The record length specified in columns 9-12 of control card 1 is greater than the maximum allowable record length.	Correct the control card and/or condition and restart the sort.
7.	4248	847	RECORD SIZE TOO SMALL - RESTART Reason: The input record length specified in columns 9-12 of control card 1 is less than the minimum allowable length (ten characters if blocked, thirteen characters if unblocked).	Correct the control card and/or condition and restart the sort.
8.	4412	876	OVERSIZE INPUT BLOCKING, GREATEST POSSIBLE B XXX - RESTART Reason: The input blocking factor specified in columns 13-15 of control card 1 is greater than the maximum possible sort blocking factor.	Correct the control card and/or condition and restart the sort.
9.	4431	882	BOL TOO LARGE - RESTART Reason: The specified output block length specified in columns 70-73 of control card 1 is greater than the sort block length.	Correct the control card and restart the sort.
10.	5111	992	BO NOT SUBMULTIPLE OF B, TO SET BO EQUAL TO XXX PRESS START Reason: The output blocking factor specified in columns 16-18 of control card 1 is not a submultiple of the program-calculated sort blocking factor. The program provides the user with the option to continue processing with the output blocking factor equal to the sort blocking factor. BO NOT SUBMULTIPLE OF B, TO SET BO EQUAL TO XXX PRESS START. TO MAKE B A MULTIPLE OF BO, PRESS START/RESET AND START Reason: The output blocking factor is not a submultiple of the sort blocking factor. The program, if the user so desires, will recalculate the sort blocking factor on the basis of the output blocking factor instead of on the basis of the input blocking factor.	Press the start key to continue processing with the output blocking factor equal to the sort blocking factor. Otherwise correct the control card and restart the sort. Press the start-reset key and then the start key if the program is to recalculate the sort blocking factor. Otherwise, press the start key to continue processing.
11.	5173	1005	BO GREATER THAN B, BO MAY EQUAL XXX OR ANY SUBMULTIPLE OF IT. PRESS START FOR BO EQUAL B Reason: The output blocking factor specified in columns 16-18 of control card 1 is greater than the program-calculated sort blocking factor. The program provides the user with the option of continuing with the output blocking factor equal to the sort blocking factor.	Press the start key to resume processing with the output blocking factor equal to the sort blocking factor. Otherwise correct the control card and restart the sort.

Assignment Phase Halts and Messages (Part 1 of 2)

[1] International Business Machines, Data Processing Division, *Sort 7 Specifications and Operating IBM 1401 and 1460,* 55p. (White Plains, N.Y., 1963).

Number	I-Address Register	Sequence Number	Message and/or Reason	Procedure
12.	5884	1122	EFFECTIVE FILE SIZE GREATER THAN MAXIMUM FILE SIZE. PRESS START TO CONTINUE. Reason: The specified file size is greater than the program-calculated maximum allowable file size.	Either press the start key to continue processing using the program-calculated file size, or correct the control card and condition and restart the sort.
13.	5980	1131	TLCF GREATER THAN L Reason: The total control-data field length specified in columns 30-32 of control card 1 is greater than the specified input record length.	Correct the control card and restart the sort.
14.	6591	1200	One of the following five messages will be associated with this halt. (1) AN INVALID CONTROL CARD IS PRESENT OR OUT OF SEQUENCE. CHECK COLUMN XX-XX. (2) TAPE UNIT IS INVALIDLY SPECIFIED. (3) RCC IS NOT SPECIFIED FOR VARIABLE (4) NO. OF CONTROL FIELDS IS INVALIDLY SPECIFIED (5) NO. OF INPUT REELS IS INVALIDLY SPECIFIED Reason: The program has determined that incorrect data has been punched in the control card. The incorrect field is shown in each message. Message (3) is printed when columns 58-61 in control card 1 are blank and variable-length blocked records are specified.	Correct the control card(s) and restart the sort.
15.	6652	1206	USER AREA SPECIFIED ABOVE CORE SIZE RESTART Reason: The starting address of the Phase-2 user area specified in columns 52-56 of control card 1 is greater than the specified core-storage capacity.	Correct the control card and restart the sort.
16.	6814	1236	URPI IS INVALIDLY SPECIFIED, PRESS START FOR PUNCH OPTION Reason: The unreadable-record procedure is invalidly specified in column 19 of control card 1.	Press the start key to select the punch option. Otherwise, correct the control card and restart the sort.
17.	6862	1247	AN INVALID CONTROL CARD IS PRESENT Reason: The message explains the reason for this halt.	Check the control cards, insert the proper control cards in the proper sequence, and restart the sort.
18.	6884	1253	CONTROL CARD #4 AND/OR #5 ARE MISSING Reason: The user has selected options that require control card 4 and/or control card 5, and the card(s) are not present.	Check the control cards, insert the proper cards in the proper sequence, and restart the sort.
19.	6953	1262	TAPE UNIT NOT AVAILABLE - PRESS START FOR PUNCH OPTION Reason: The user has specified that any unreadable tape blocks are to be written on tape and Phase 2 is to be a three-way merge.	Either press the start key if unreadable tape blocks are to be punched into cards or correct the control card and restart the sort.
20.	7009	1274	THE LENGTH OF RCC IS NOT SPECIFIED FOR VARIABLE Reason: The user has specified that variable-length records are to be read in the load mode and that each record has a record-character-count field. However, column 1 of control card 3 does not contain a 3 or 4.	Correct the control card and restart the sort.

Assignment Phase Halts and Messages (Part 2 of 2)

Number	I-Address Register†	Sequence Number †	Message and/or Reason	Procedure
1.	782	3257	A message is not associated with this halt. Reason: A redundant input block or a wrong-length fixed-length record has been read and printed when the print-scan or correct option is specified.	For redundant input blocks, follow the procedure described under Print-Scan or Correct Option. For wrong-length records, press the start key to delete the record.
2.	803	3294	A message is not associated with this halt. Reason: End-of-reel has been reached on the tape on which any unreadable input tape blocks are being written.	Rewind and unload the tape. Mount a new tape and press the start key.
3.	1796	1654	A message is not associated with this halt. Reason: A write redundancy indication remains after twenty attempts to write a specific block on the output tape. The tape-select light of the unit on which the write redundancy occurred is ON.	Either press the start key to have the program try to write the block twenty more times or mount a new tape. (Refer to Write Redundancy)
4.	1807	1656	A message is not associated with this halt. Reason: The number of records in the input file is greater than the maximum allowable number.	Either press the start key if it is known that the output tape can contain the input file, or correct the condition and restart the sort.
5.	1848	3408	EOR ON 2 OUTPUT TAPES Reason: The input file cannot be written on the two output tape reels. This condition can result if the tapes are not full length.	Either decrease the file size or mount two full reels of tape and restart the sort.
6.	1933	1689	A message is not associated with this halt. Reason: Fixed-length records are being processed and the program has read a wrong-length record.	Press the start key to handle through the redundancy routine. The record will be written on tape, punched, or printed depending on the specification in column 19 of control card 1. If the record is printed, halt number 1 occurs.
7.	2235	1752	A message is not associated with this halt. Reason: End-of-file has been reached on an input tape and sense switch B is ON.	Load the next input tape if another tape is to be processed. Press the start key to continue processing.
8.	2565 (B) 2566 (A)	1816 (B) 2288 (A)	ERROR - SHOULD BE (80 character label) Reason: An error was detected in checking the input header label. The input header label is printed before this message prints. The label printed in this message contains the correct input header label information.	Press the start key if the label is acceptable as read. To recheck the label or, if a new tape has been mounted, press the start-reset key and then the start key to continue processing.
9.	2735 (B) 2713 (A)	1848 (B) 2317 (A)	ERROR - SHOULD BE (input block count kept by the Sort-7 Program) Reason: The input block count in the input trailer label being processed does not equal the block count kept by the program. The trailer label is printed before this message.	Either press the start key to continue processing if the block count kept by the program is correct, or correct the error condition and restart the sort.
10.	2739 (B) 2717 (A)	1849 (B) 2318 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read an input trailer label. The input error options are not available when the program encounters an unreadable trailer label.	Press the start key to have the program ignore the label and continue processing.
11.	4564 (B) 4562 (A)	2121 (B) 2375 (A)	RETAIN TAPE X FOR XXX DAYS Reason: The retention cycle of the tape on tape unit X indicates that this tape should be retained for XXX days.	Press the start key to continue processing if the tape can be used for output by Sort 7. If a new tape with a header label is mounted, press the start-reset key and then the start key to continue processing.

† (A) indicates that 120-character labels (Type A) are being processed.
(B) indicates that 80-character labels (Type B) are being processed.

Phase 1 Halts and Messages (Part 1 of 2)

Number	I-Address Register	Sequence Number	Message and/or Reason	Procedure
12.	4600(B) 4598 (A)	2129 (B) 2383 (A)	A message is not associated with this halt. Reason: Retention-cycle checking routine has exceeded a 30-year period.	Press the start key to accept tape and continue processing. Press the start-reset key and start key to retry retention test.
13.	4869 (B) 4854 (A)	2192 (B) 2443 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read a header label on a work tape.	Either press the start-reset key and then the start key to have the program attempt to read the label one hundred more times, or press the start key to accept the label as read, or mount a new tape with a header label and press the start-reset key and then the start key to continue processing.
14.	4987 (B) 4972 (A)	2214 (B) 2465 (A)	A message is not associated with this halt. Reason: A write-redundancy indication remains after twenty attempts to write a header label on a work tape.	Either press the start key to have the program attempt to write the label twenty more times, or mount a new tape and restart the sort.
15.	5153	3181	A message is not associated with this halt. Reason: A write-redundancy indication remains after twenty attempts to write a checkpoint record.	Either press the start key to have the program attempt to write the record twenty more times, or mount a new tape and restart the sort.
16.	5157	3182	A message is not associated with this halt. Reason: End-of-reel has been reached while the program was attempting to write a checkpoint record.	Press the start key to continue processing. The program attempts to write the checkpoint record on the tape.
17.	5264	3204	PHASE 1 INP X X X OUT X X X Reason: A Phase-1 restart has been performed. The input tapes and the output tapes must be mounted on the specified tape units.	Press the start key to resume processing after the tapes have been properly mounted and assigned and the program deck starting with card 598 has been placed in the card reader.
18.	5275	2654	PATCH PROGRAM TOO LARGE PRESS START - IGNORE PATCH Reason: The area of core storage that is reserved for a user-written routine is too large and the Phase-1 internal processing cannot be performed. The user is given the option of restarting or of continuing the program with the user routine being deleted.	Either press the start key to continue processing without the user routine, or correct the condition and restart the sort.
19.	5688	2734	OVERSIZE INPUT BLOCKING - RESTART Reason: The specified input block length is greater than the maximum allowable length.	Correct the control card and/or condition and restart the sort.

Phase 1 Halts and Messages (Part 2 of 2)

Number	I-Address Register	Sequence Number	Message and/or Reason	Procedure
1.	611	3519 (B) 3732 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read a header label on a Phase 2, pass 1, output tape.	Press the start key to accept the label as read and to bypass the retention cycle check, or press the start-reset key and then the start key to have the program attempt to reread the label 100 more times.
2.	619	3530 (B) 3742 (A)	A message is not associated with this halt. Reason: A write-redundancy indication remains after twenty attempts to write a header label on a Phase 2, pass 1, output tape.	Press the start key to have the program use this tape as a work tape, or mount a new tape and press the start-reset key and then the start key. Exercising the first option may cause a halt during the next Phase-2 pass.
3.	634	3540 (B) 3750 (A)	RETAIN TAPE X FOR XX DAYS Reason: The retention cycle in a header label written on a Phase 2, pass 1, output tape indicates that the tape should be saved.	Press the start key to have the program write a header label on this tape and continue processing. If a new tape is mounted, press the start-reset key and then the start key to continue processing.
4.	642	3549 (B) 3758 (A)	A message is not associated with this halt. Reason: A tape mark is not written after a header label when it is specified that a tape mark follows each header label.	Insure that either the control cards are correct or that all header labels have tape marks following them, and then restart the sort.
5.	646	3553 (B) 3762 (A)	A message is not associated with this halt. Reason: A Phase 2, pass 1, input tape contains an unreadable header label.	Either press the start key to bypass the label, or press the start-reset key and then the start key to have the program attempt to read the label 100 more times.
6.	863 (B) 884 (A)	3574 (B) 3783 (A)	INPUT TAPES MAY BE SAVED Reason: A sense-switch-D interrupt has occurred. The user can unload the Phase-1 input tapes and mount work tapes on the input tape units.	Mount the required work tapes and press the start key to continue processing.
7.	1267 (B) 1279 (A)	3651 (B) 3857 (A)	Amessage is not associated with this halt. Reason: Retention-cycle checking routine has exceeded a 30-year period.	Press the start key to accept tape and continue processing. Press the start-reset key and the start key to retry retention test.

Prephase Label Handling Routine Halts and Messages

Number	I-Address Register	Sequence Number†	Message and/or Reason	Procedure
1.	710	3970	A message is not associated with this halt. Reason: A read-redundancy occurred during the reading of a checkpoint at the end of a phase 2 merge pass.	Press the start key to continue processing using the checkpoint record as read. Press the start-reset key to have the program retry to read the checkpoint record.
2.	2549	4307	A message is not associated with this halt. Reason: End-of-file has been reached during the reading of the first input tape of the current merge pass.	Replace the tape on which the end-of-file indicator was detected and then restart the sort.
3.	2622	4325	A message is not associated with this halt. Reason: A write-redundancy has occurred on an output tape during the current merge pass. Twenty attempts to write the output block have been made. The tape-select light of the unit on which the write redundancy occurred is ON.	Press the start key to attempt to write the block twenty more times.
4.	3031	4895	A message is not associated with this halt. Reason: A redundant tape block has been read and printed when the print-scan or correct option is specified.	Follow the procedure described in Print-Scan or Correct Option.
5.	3038	4856	A message is not associated with this halt. Reason: End-of-reel has been reached on the tape upon which unreadable tape blocks are being written.	Rewind and unload the dump tape, mount a new tape, and press the start key.
6.	3091	4401 (B) 4716 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read a header label.	Press the start key to accept the label as read. Press the start-reset key and then the start key to have the program try to read the label 100 more times.
7.	3091	7898 (B) 8110 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read the header label on the final output tape.	Press the start key to accept the label as read. The retention cycle will not be checked. Press the start-reset key and then the start key to have the program attempt to read the label one hundred more times. The option of mounting a new tape is also provided. In this case, mount the tape, press the start-reset key and then the start key to continue processing.
8.	3096	4407 (B) 4721 (A)	A message is not associated with this halt. Reason: A write redundancy has occurred during the writing of a header label. Twenty attempts to write the record have been made.	Press the start key to have the program attempt to write the header label twenty more times.
9.	3099	7905 (B) 8115 (A)	A message is not associated with this halt. Reason: A write-redundancy indication remains after twenty attempts have been made to write the header label on the final output tape.	Press the start-reset key and then the start key to attempt to write the record twenty additional times. The option of mounting a new tape is also provided. In this case, mount the tape and then press the start key to continue processing.
10.	3111	4411 (B) 4724 (A)	A message is not associated with this halt. Reason: A read-redundancy indication remains after one hundred attempts to read a header label.	Press the start key to accept the label as read. Press the start-reset key and then the start key to have the program try to read the label 100 more times.

† Same as Figure 15, Part 1.

Phase 2 (Balanced Merge) Halts and Messages (Part 1 of 3)

Number	I-Address Register	Sequence Number †	Message and/or Reason	Procedure
11.	3114	7909 (B) 8118 (A)	RETAIN TAPE X FOR XXX DAYS Reason: The retention cycle of the tape on tape unit X indicates that the tape should be retained.	Press the start key if the tape is to be used. If a new tape is mounted, press the start-reset key and then the start key to continue processing.
12.	3122	7913 (B) 8121 (A)	A message is not associated with this halt. Reason: A write-redundancy indication remains after twenty attempts have been made to write the trailer label on the final output tape.	Press the start key to accept the label as written. Press the start-reset key and then the start key to attempt to write the label twenty additional times.
13.	3126	4416 (B) 4727 (A)	A message is not associated with this halt. Reason: A tape mark is not written after a header label when it is specified that a tape mark follows each header label.	Insure that either the control cards are correct or that all header labels have tape marks following them, and then restart the sort.
14.	3128	5610	A message is not associated with this halt. Reason: A three-way merge is being performed, tape labels are not being processed, two output tapes have become full during the final merge pass and additional output records remain to be written.	Mount new tapes on the last two output tape units and press the start key.
15.	3132	5611	A message is not associated with this halt. Reason: A two-way merge is being performed, tape labels are not being processed, the output tape has become full during the final merge pass, and additional output records remain to be written. The tape has been tape marked, rewound, and unloaded.	Mount a new tape on the tape unit on which the tape-select light is ON. Press the start key to continue processing.
16.	3137	7918 (B) 8124 (A)	(B) CHG TP X (A) CHANGE REEL ON UNIT X Reason: End-of-reel has been reached on the final output tape that is mounted on unit X.	Mount a new tape reel and press the start key to continue processing.
17.	4477	4954	A message is not associated with this halt. Reason: There is an error in the program calculation of the number of merge passes required in Phase 2.	Decrease the size of the input file and restart the sort.
18.	4801	5265	XXXXXX REC PROCESSED - PASS XX ERROR Reason: The record count kept by the program during the previous merge pass does not equal the record count kept by the program during the current merge pass.	Press the start key to continue processing, or restart the pass as described in Checkpoint, Interrupt, and Restart.
19.	4934	5291	LAST MERGE PASS INTERRUPT Reason: All Phase-2 merge passes except the final pass have been completed. This halt occurs before the initialization of the final pass.	Press the start key to continue processing.
20.	5674	5405	DENSITY MAY BE CHANGED Reason: Sense switch D is ON. The halt occurs before the checkpoint record is written on the first output tape.	Press the start key to continue processing.
21.	5936	5456	PASS XX - CKPT X - INP X, X, X, INTERRUPT (HIGH) (LOW) DENSITY Reason: Sense switch E is ON. This halt occurs after the checkpoint record for the next merge pass has been written. The message shows the number of the merge pass just completed, the tape on which the checkpoint record is written, and the tapes that contain the input for the next merge pass.	Press the start key to continue processing, or interrupt the program and restart processing at a later time. In the latter case, save the input tapes and the tape containing the checkpoint record.
22.	6040	5477	A message is not associated with this halt. Reason: A Phase-2 restart has been performed using the systems tape. The program is ready to position the systems tape.	Insure that all tapes are correctly mounted and addressed. Press the start key to continue processing.

† Same as Figure 15, Part 1.

Phase 2 (Balanced Merge) Halts and Messages (Part 2 of 3)

Number	I-Address Register	Sequence Number	Message and/or Reason	Procedure
23.	6102	5488	PLACE CARDS FROM 1468 IN READER PASS XX INP X, X, X, OUT X, X, X, (HIGH) (LOW) DENSITY. Reason: A Phase-2 restart has been performed after a sense-switch-E interrupt. The message shows how many merge passes have been completed, the numbers of the input and output tape units, and the density that was specified in control card 1.	Refer to Restart for the procedure to be followed.
24.	6274	5515	A message is not associated with this halt. Reason: End-of-file occurred while bypassing the old checkpoint record on restart.	Press the start key to accept the record as read and continue processing.
25.	6299	5520	A message is not associated with this halt. Reason: 99 attempts to read a checkpoint failed.	Press start to retry the read operation. If the read error persists, restart the entire program.
26.	6319	5524	A message is not associated with this halt. Reason: End-of-file occurred while attempting to read the checkpoint record.	Press the start key to continue processing. Press the start-reset key and then the start key to backspace the tape and reread the checkpoint record.
27.	6327	5526	OUTP X - END OF SORT Reason: The sort run is completed and the final output tape on unit X is rewound and unloaded.	
28.	6579	5567	A message is not associated with this halt. Reason: A write-redundancy has occurred while writing the checkpoint record. Twenty attempts to write the record have been made.	Press the start key to have the program attempt to write the record twenty more times.
29.	6583	5568	A message is not associated with this halt. Reason: End-of-reel has occurred while the program was attempting to write a checkpoint record.	Press the start key to accept the checkpoint record as written. Otherwise, restart the sort run.

Phase 2 (Balanced Merge) Halts and Messages (Part 3 of 3)

ERROR LISTINGS AND SYSTEMS MESSAGES

Thesaurus Module

XMAIN0 – Pre-List of the Thesaurus Input

Terms not in alphabetical sequence. The input terms must be in sequence by accession number, columns 73-76.

First card of set not '1', '2', or '3' in col. 80. The only possible entry codes are '1', '2', '3', 'J', 'K', or 'L'.

Term code not '', '#* or '-' col. 1. Each descriptor, (entry code '1', '2', 'J' or 'K') must contain a pre-code. The only permissible values for pre-code are '*', '#' or '-'. Either pre-code has been omitted or an invalid entry specified.

Cards out of sequence on col. 77. All cards within a term set must be in ascending alphabetical sequence on column 77.

Substitute pre-code not '', '#* or '-' col. 1. When a substitute has been specified for a descriptor, it must also be assigned a pre-code. The only permissible entries are '*', '#' or '-'. Either pre-code has been omitted or an invalid entry specified.

More than 16 cards with 4 col. 80. A term set may not contain more than sixteen units of cross reference or scope note information. If more than sixteen cards are to be entered, a continuation record must be created. (See Chapter VII, specifically Thesaurus Data Entry.)

XMAIN1 – Format the Thesaurus Input

T01 – First unit of a set not term card. The first unit of a set must be the term card which is identified by a blank in column 77.

T02 – Invalid entry code in term card. The only possible entry codes for a term transaction are '1', '2' or '3' (for the addition of a term) or 'J', 'K', 'L' (for the deletion of a term) in column 80.

T03 – Invalid type code or blank term entry. A descriptor must be assigned a pre-code for search purposes. The only permissible entries are '*', '#' or '-'. The term either contained an invalid entry code or the term field on the card was left blank.

T04 – Truncation feature not used properly. Either column 79 contains a '1' and no truncation symbol '$' was specified in the body of the term (columns 1-54) or a truncation symbol was specified but column 79 did not contain a '1'.

T05 – Substitute pre-code invalid. When a substitute has been specified for a descriptor, it must also be assigned a pre-code. The only permissible entries are '*', '#' or '-'. Either pre-code has been omitted or an invalid entry specified.

T06 – Units out of sequence. All the units within a set must be ascending sequence according to column 77.

T07 – Invalid numerical term. A numerical term (first character numeric) may not contain embedded blanks.

T08 – Too many cross-references. A term set may not contain more than sixteen units of cross-reference or scope note information. If more than sixteen units are to be entered, a continuation record must be created. (See Chapter VII, specifically Thesaurus Data Entry.)

XXX – Unedited unit. A previous unit within the term set contained an error. All following units are therefore rejected. This unit was not edited and has been removed from further processing since the entire set will have to be resubmitted.

XMAIN2 – Thesaurus File Update

T51 – Invalid deletion of a term. An attempt has been made to delete a term that is not on the Thesaurus File.

T52 – Duplicate term. An attempt has been made to add a term that is already on the Thesaurus File. Duplicate terms are not allowed.

T53 – Duplicate truncation root. A truncated term already appears on the Thesaurus File with a portion or the whole root of a term. Duplicate term roots are not allowed.

T54 – Duplicate input term. The same term appears more than once in the input to the Thesaurus File. All occurrences after the first have not been processed. Duplicate terms are not allowed.

T55 – An input term duplicates an existing term root. A term root already exists on the Thesaurus File. The input term contains the term root as the stem of the term. Duplicate terms and term roots are not allowed.

T56 – An input term contains the term root of an existing term. An existing term on the Thesaurus File contains the root of the truncated term being entered. Duplicate terms and term roots are not allowed.

DEL – Term deleted. This message is for audit trail purposes. It is an indication that a term has been either deleted entirely from the Thesaurus File or has been temporarily removed so that a new version may be added.

XMAIN5 – Permuted Thesaurus

If any lines are printed while the cards are being read, it indicates that a word of more than twenty characters appears in a term. If this happens, that permutation of the term is eliminated from the Permuted Thesaurus.

Pre-edit Module

E0020 – Error Listing

More than ten '22's. E0020 can generate a maximum of '22' units per document reference. This message indicates that there are more than ten such units. All '22' units after the tenth have not been processed.

Too many headings. E0020 can generate a maximum of ten subject headings, '99' units, per reference. This message indicates that there are more than ten such units. All '99' units after the tenth have not been processed.

E0030 – Pre-Edit Listing of the Document Input

01 – No main author. Each document reference must contain a personal, corporate or anonymous author, coded as '01', '02' or '03' respectively.

02 – No title. Each document reference must contain a title entry which can be either a significant title '11' or a non-distinctive title '15'. Titles can also be noted in elements '16', '17' and '18', elements allocated for foreign languages titles in French, German and Spanish respectively. If a title other than an '11' or a '15' has been selected, an error will be noted on the error legend and must be verified to determine whether or not one of the other types of title entries has been assigned.

03 – No imprint or collation. Each document must contain an element '21' which provides imprint information.

04 – No subject heading. Each document must have at least one descriptor which has been assigned as a subject heading ('SH' noted in columns 21-22) for which a '99' entry is to be generated.

05 – Document number out of sequence. All documents must be in sequence by document number.

06 – First record not type '1'. The first unit of each document must contain a heading card, '1' in column 20, which provides the entry date of the document into the URBANDOC system.

07 – Unit number out of sequence. The unit within the document are arranged in numeric sequence. If two units are out of sequence or unit numbers are identical, this error is noted.

08 – Entry not 'A' or 'D'. Each unit must contain either an 'A' or 'D' in column 19.

09 – Record type not '1', '2', '3' or '5'. Column 20 of each unit must contain one of the following codes: '1', '2', '3' or '5'. The '1' is used for the very first record in the document, the '2' for content analysis data, the '3' for subdescriptors within the content analysis data, and the '5' for the free text or descriptive analysis data. Any other value in column 20 is considered an error.

10 – More than one type '1' record. Only the very first record within the document, the heading unit, can contain a '1' in column 20.

11 – Unit number greater than '1000'. This message applies to the unit numbers for content analysis records. These records cannot have a unit number higher than '0999'.

12 – Element number not blank. This error message also applies to the content analysis section of the document record. Columns 21-22 must be blank for descriptors.

13 – Pre-code not '', '#' or '.'*. This message applies only to descriptors, both subject and geographic, which must contain either '*', the pre-code for most subject descriptors, the ' ', the pre-code for common subject descriptors, or the '-', the pre-code for all geographic descriptors. The message indicates an invalid pre-code in column 23.

14 – Subdescriptor longer than 12 characters. The first subdescriptor assigned to a document is in columns 59-70. It may not be longer than twelve characters.

15 – Column 24 not '-'. This error message applies only to place name descriptors which must have a '-' block code in column 24. This '-' distinguishes the geographic descriptors from the subject descriptors.

16 – Last character not a '.', '' or '/'*. This error message applies only to those place name descriptors which relate to code forms. Except for authorized exceptions, each of these terms must end in either '.', '*' or '/'. The document analyst must determine if it is an error or an authorized exception by consulting the Thesaurus.

17 – Element number not 'XX'. This message pertains to units with more than one subdescriptor per descriptor. These units must contain 'XX' in columns 21-22.

18 – Pre-code not 'N'. The second and following subdescriptors for a descriptor must contain an 'N' in column 23.

19 – 1st six characters of the descriptor incorrect. When there is more than one subdescriptor per descriptor, the descriptor being modified must have its first six characters repeated beginning in column 25.

20 – Subdescriptor longer than 12 characters. The second and following subdescriptors for a descriptor are in column 35-46. It may not be longer than twelve characters.

21 - Invalid element number. This error message applies only to the descriptive analysis section of the record. An invalid element number has been specified in columns 21-22.

22 – Unit out of sequence by eelement. This error message pertains to the descriptive analysis portion of the reference. All bibliographic information must be in sequence by element number, columns 21-22.

E0040 – Update the Document Input File

ADD – Addition to file. A unit has been added to the Document Input File.

DEL – Deletion from file. A unit has been deleted from the Document Input File.

CHG – Change to file. A change has been made to a unit already on the Document Input File.

SEQ – Type '5' record out of sequence. Within the same document number, a descriptive analysis record may be only followed by another descriptive analysis record or the beginning of a new document reference. Descriptors or subdescriptors cannot appear within or after the descriptive analysis information for a document.

ERR – Non-processed record. Entry code is not an addition, deletion or change, or an attempt has been made to delete a record that does not appear on the file.

File Maintenance Module

MAINT1 – CFS Format of the Document Input File

Out of sequence. The input units must be in sequence by document number and unit number. The output unit was out of sequence and has been removed from further processing.

MAINT3 – Edit Listing

X01 – Invalid term. The descriptor or subdescriptor used in this unit is not included on the Thesaurus File and has been rejected from further processing.

X02 – Incorrect pre-code. Each descriptor on the Thesaurus File has an authorized pre-code of '*', '#' or '-'. The pre-code for the descriptor in the input does not agree with the pre-code assigned to the term on the Thesaurus File.

X03 – Incorrect usage of term. Each term on the Thesaurus File is coded for use as either a descriptor, subdescriptor, or both. The usage of this term on the input does not agree with the code on the Thesaurus File.

X99 – Invalid subject heading. Each term on the Thesaurus File may be used as either a descriptor or a subject heading. The term in the input unit has been used as a subject heading and is not included on the Thesaurus File.

XXX – Substitute term used. Each term on the Thesaurus File has specified forms for a concept. The original term entered in the input unit has been replaced with the preferred form shown on the listing. It is this form that is entered onto the Document Master File.

003 – No descriptor units. Each reference added to the Document Master File must contain at least one descriptor. Either all the descriptors assigned were invalid or the heading unit, '1' in column 20 indicating that a new document is to be added, has been separated from its descriptors.

004 – Unit number out of sequence. Each unit of a document must appear in ascending order on columns 15-18.

005 – Heading unit incorrect. The heading unit, '1' in column 20, marks the addition of a new document reference to the Document Master File. This must be the first unit of the document. The heading unit is either duplicated or specified incorrectly.

006 – Deletion incorrect. An attempt has been made to delete a bibliographic element from a reference on the Document Master File. The element for that reference does not appear on the file.

008 – Subdescriptor associated with wrong descriptor. Each subdescriptor after the first subdescriptor, '3' units, must immediately follow its parent descriptor, '2' units, and must contain in columns 25-30 the first six characters, excluding the pre-code, of its descriptor. The descriptor and its subdescriptor have been separated or the descriptor or the first six characters of its root have been misspelled.

011 – Invalid entry code. The valid entry codes for column 19 are 'A' or 'D'. For column 20 the valid codes are '1', '2', '3' or '5'. The contents of either or both columns are invalid.

015 – Columns 21-22 not blank. Columns 21-22 must be blank on a heading unit, '1' in column 20. This field is not blank.

016 – Document number out of sequence. The document numbers must be in ascending order by document number, columns 1-14.

017 – Element number out of sequence. The descriptive analysis information must be in sequence by element number, columns 21-22.

018 – Invalid element number. The valid element numbers range from '01'-99'. Element number, columns 21-22, contains something other than these numbers.

019 – Deletion of descriptor and subdescriptor incorrect. Deleting a descriptor, 'D2' in columns 19-20, from the Document Master File automatically deletes all subdescriptors for that descriptor. Therefore, the specification to delete a descriptor cannot include a subdescriptor. Columns 47-80 were not blank.

023 – Values present for duplicate subdescriptors. A value has been included for a subdescriptor which duplicates a term already assigned to the descriptor. The value will not be processed.

024 – Subdescriptor addition or deletion incorrect. To add or delete a subdescriptor to a descriptor in a document record, the input unit must include the descriptor's rank (or number) and first six characters, excluding the pre-code. Columns 25-30 in the input unit do not agree with the first six characters of the descriptor with that rank.

025 – Columns 23-24 not 'N'. The contents of columns 23-24 must be 'N' for a subdescriptor being added to or deleted from the Document Master File record.

028 – Placement number not specified. To delete a subdescriptor from the Document

Master File, the input must specify the descriptor's rank and first six characters, excluding pre-code, the subdescriptor to be deleted and its placement number. The placement number on this input unit was not specified.

030 – Subdescriptor too long or code incorrect. When this error is associated with an input unit other than for date analysis, it indicates that the subdescriptor has exceeded the maximum permissible length of twelve characters. When associated with an input unit for date analysis, it indicates a miscoding of the type of date. The possible codes are 'S' for single date, 'U' for an upper range of a date, and 'L' for a lower range of a date.

031 – Error in range option for date analysis. When specifying a range, both the upper and lower limits must be present on the same input unit. This input unit does not contain both limits.

032 – Duplicate subdescriptors for one descriptor. The same subdescriptor may not be entered more than once for any one descriptor.

033 – Invalid pre-code. The only valid pre-codes for a descriptor are '*', '#' or '-'. One of these characters does not appear in column 23 of the input unit.

034 – Unit bypassed and unedited. An error in a preceding input unit has caused this unit to be bypassed and not edited.

037 – Heading unit last unit of a document set. When a heading unit appears as the last unit of a document set, it is treated exactly as if it had no descriptors associated with it. (See 003.)

Z34 – Subject heading bypassed due to previous error. This subject heading could not be included in the Document Master File, not because it is an invalid term, but because of certain preceding conditions. (See 034.)

MAINT6 – Non-Processed Units

002 – Oversize searchable segment. The descriptor and/or subdescriptor addition was not processed. Its addition to the reference would have caused that segment to exceed the 2200 character limitation.

003 – Oversize free text segment. The bibliographic unit was not processed. Its addition to the reference would cause that segment to exceed the 2200 character limitation.

004 – Invalid modifier addition. An attempt has been made to add a modifier to a descriptor in an existing document reference. The modifier was not added because the parent descriptor could not be located in the record.

005 – Duplicate document number. The new document reference was not added to the file since a reference already exists with that document number. The original reference with that document number contains a mixture of the original entry and some data from

the duplicate entry. The original reference must be deleted and the duplication of identification numbers resolved. Both references must be re-entered.

006 – Non-existent record. An attempt has been made to revise a reference on the file. The revision was not made because the reference with that document number could not be located.

009 – No searchable section. An attempt has been made to create a document reference for which there would have been no searchable section.

012 – Edit input out of sequence. The Edit File is out of sequence by document number and unit number. This error should not occur. If it does, it can probably be corrected by resorting the Edit File.

013 – Invalid free text addition. An attempt has been made to delete or change one free text segment (bibliographic element). The specified record could not be located.

017 – Invalid descriptor root on addition. A modifier to be added to a descriptor in an existing document reference will specify both the rank and the first six characters of the parent descriptor. The descriptor root in the unit does not match the first six characters of the descriptor on the Document Master File with the specified rank.

021 – Invalid modifier deletion. An attempt has been made to delete a modifier from a descriptor on the Document Master File. The parent descriptor specified in the input unit could not be located in the reference.

022 – Invalid document addition. An entire document reference has been removed from the Document Master File. An attempt to reenter or revise that reference (no longer on the file) cannot be made without a new heading unit.

023 – Non-processed descriptor deletion. An attempt has been made to delete the last remaining descriptor from a document reference. The change has not been made since it would have the effect of deleting the entire reference and a "delete reference" was not specified.

024 – Invalid descriptor root on deletion. A modifier to be deleted from a descriptor in an existing document reference will specify both the rank and the first six characters of the parent descriptor. The descriptor root in the input does not match the first six characters of the descriptor (on the Document Master File) with the specified rank.

025 – Duplicate modifier. An attempt has been made to add a duplicate modifier to a descriptor in a document reference.

027 – Duplicate descriptor. An attempt has been made to add a duplicate descriptor to a document reference. This occurs most frequently for the geographic analysis terms when many place names substitute to the same city size and SMSA code. For this case, no action is necessary.

MAINT7 – Merge Two Document Master File

702 – Oversize document records. An oversize document record has been encountered. This record has caused the job to be ended. A memory dump at the time of job discontinuance is available.

705 – Duplicate document numbers. When merging two files, one is considered "prior", the other an "update". The reference on the "update" file is a duplicate of the reference number on the "prior" file. The "update" reference is treated as the error and will not be included on the output file.

709 – Invalid document reference addition. The document reference on the "update" file does not contain a searchable section. This reference will not be included on the output file.

710 – Invalid document reference on prior. The document reference on the "prior" file does not contain a searchable section. This reference will not be included on the output file.

MAINT8 – Statistical Analysis of the Document Master File

This program checks the documents for authorized element or segment numbers. If an unauthorized element number is present, the document number, element number, and the first line of the bibliographic information of the unauthorized record is printed. This listing is designed to allow the document analyst to be aware of variances over time in the use of bibliographic elements. This printed information would be useful if the bibliographic contents of the Document Master File were to be made consistent.

Search Module

SEP1 – Expand the Requests

Col. 80 not blank. The logical expression is limited to a maximum length of seventy-three characters. If column 80 is not blank, either the request equation is too long or it has been written incorrectly. The request will not be processed.

Unmatched parentheses. The system allows the user to nest expressions to be expanded by the system. These expansions are indicated by parentheses; however, there must be an equal number of left and right parentheses. A request equation without an equal number of left and right parentheses will not be processed.

Request wrong. The logical expression contains embedded blanks. The request will not be processed.

Exceeds CFS search limitations. The logical expression is entered as a seventy-three character expression which may be expanded by the system. Examination of the logical expression after expansion reveals that it will exceed the limitation on the number of characters in a request. The request will be rejected.

Wrong request number. All the units for a single request must have the same request number in columns 1-2. Request numbers must be in ascending order. This error will cause the search to be cancelled.

Wrong number of title (and/or) format cards. There may be two title cards and one format card per request set. There must be at least one title card and one format card per set. If a request does not meet these specifications, the search will be ended.

Company out of sequence. The field "Company" is the field "Requestor Identification" in columns 6-9. (See Chapter VII, specifically Search Data Entry.) The request must be in sequence by this field. A strict accession number may be assigned to satisfy the requirement. This error will cause the search to be cancelled. This field could be used to identify an individual, group, division, etc.

Profile out of sequence. The field "Profile" is the field "Number in Requestor Set" in columns 12-16. Within "Company" or "Requestor Identification", the request must be in sequence by this field. A strict accession number may be assigned to satisfy the requirement. This error will cause the search to be cancelled. If "Company" were used to identify a group or division, "Profile" could identify an individual.

SEP2 – Check for Term Usage

Request term not found on the descriptor file. The frequency of usage will be attached to each term in a request. If a term has been used as a descriptor, it will appear on the Inverted File. If a term cannot be found on the Inverted File, it has not been used in document analysis. The user should examine the request containing this term to determine if that request should be eliminated from the search batch. This should certainly be the case in which the presence of the term is mandatory. For a request in which the term is optional (an *OR* usage), the search may be processed and still yield retrievals.

SEP3 – Format the Requests for CFS

Insufficient term table. Each symbol in the logical expression must correspond to an entry in the term table. A symbol in the expression did not have a corresponding term entry in the table. This error will cause the search to be cancelled.

Extra terms in the term table. Each symbol in the logical expression should correspond to an entry in the term table. An entry in the term table was not used in the expression. The unused entries will be considered as extra terms. The request will be processed ignoring these extra terms.

Request too long. The formatted request exceeds the CFS limitation of a search request that can be contained within thirty individual request units. This error will cause the search to be cancelled.

PHASE1 – Edit the Requests

Invalid card codes. The type for search units must be either 'T', 'F' or 'R' in column 5.

No parentheses permitted in a document search. A 'D' was specified in the Format Unit as the type of search to be performed. Somewhere in the 'R' Units (Request Units), a parenthesis appeared.

Last character of a Boolean request must be a right parenthesis. A 'B' was specified in the Format Unit as the type of search to be performed. For a Descriptor or Boolean Search, the request must end in a right parenthesis.

Unequal number of left and right parentheses. Each parenthesis marks off a set of logically grouped information. A left parenthesis marks the beginning, a right parenthesis its end. If there are not pairs of parentheses for a Boolean or a Mixed Search, the program is unable to analyze the information and perform the proper logical search.

Boolean Search must begin with a left parenthesis. The search type in the Format Unit was specified as a Boolean Search. This type of search must begin with a left parenthesis in the first 'R' unit.

Mixed Search may not begin with a parenthesis. A search type has been specified as Mixed by 'M' in the Format Unit. The proper format for the body of a Mixed Search is a list of document numbers, separated by commas and with no commas after the last document number, followed by the Boolean section of the request. The Boolean section must begin with a left parenthesis and end with a right parenthesis and follow all rules for a Boolean request.

No Title Cards. In a request, at least one Title Unit must be present. The first unit was not coded 'T' in column 5.

No Format Card. In a request, the Request Unit must be present. The card following the Title Unit(s) was not coded 'F' in column 5.

No Request Card. In the set of cards for a request, at least one Request Unit must be present. The unit following the Format Unit was not coded 'R' in column 5.

Card sequence error. In a request, the cards were not in ascending sequence by column 3-4.

More than one Format Card. In the set of cards for a request, there was more than one unit coded 'F' in column 5.

Error in Format Card. In the unit coded 'F' in column 5 there appeared at least one of the following mistakes:

There were not at least three commas in the Format Unit;

Search type was other than M(ixed), D(ocument), or B(oolean);

The option to print searchable section was answered other than Y(es) or N(o).

Maximum length of request is 20 'R' cards. The request of the search parameters (both term and logical operators) may be contained on a variable number of input units. The maximum number of request units that may be submitted is twenty. For this request, more than twenty 'R' cards were submitted.

Request has a bad term. A descriptor may not be longer than twenty-three characters. One of the terms in the request did not contain a blank after the twenty-third character. Or

For a truncated descriptor, the truncation symbol must be the last character of the term. One of the terms in the request contained the truncation symbol '$' and was not followed by a blank. Or

A subdescriptor may not be longer than twelve characters. One of the subdescriptors in the request did not contain a blank after the twelfth character. Or

A subdescriptor must be either the last word of a request or it must be followed by another search operator. One of the subdescriptors in the request was followed by something other than another operator. Or

A search operator has specified the presence of a subdescriptor and one did not immediately follow the operator. Or

No terms were specified as part of the search request.

Errors That May Be detected During the Search Process

No output from PHASE2. During the inverted file search, some terms from each request are matched against the Inverted File for a list of potential hits. This message indicated that none of the selected terms have been used during document analysis. It is most unlikely that this message will occur.

Request term is redundant. During the inverted file search, a check is made for a physical tape read error. If the read error cannot be bypassed, the term contained in the tape record is listed and non-processed. It is most unlikely that this message will occur.

No output from PHASE3. After the inverted file search, the list of potential hits will be matched and merged. The reduced list will contain only documents qualifying for each request term selected for the inverted file search. This message indicates that there are no candidates for further search from the entire batch. It is most unlikely that this message will occur.

Request and term no match. This is a systems check to be sure that each retrieval term has a full request to match. It is most unlikely that this message will occur.

Document not present. During the serial file search, the list of hits are matched to the Document Master File. This message indicates that a document reference to be searched serially is not present on the Document Master File. This condition will only occur when the Inverted File and the Document Master File are out of phase and is usually caused by an operator error. This error should not occur.

No hits located for this request. This message will be printed as part of the *Retrieval Report* when no reference meets a request's specifications.

Publication Module

P0010 – Error Listing

ERR – No match for error cards. The program matches the document number and unit number on the error card to the document number and unit number of the tape record. There is no record on tape with the document number and unit number specified on the error card.

DEL – Indication of unit deletion. When the document number and unit number on the error card match a record on tape, the tape record is deleted. This message indicates which units were deleted from the tape.

LEN – Too many text lines. With the exception of element '23', only four text lines of any one bibliographic entry are used to create a publication record. This message indicates that there were more than four lines. All lines after the fourth are eliminated.

P0040 – Excess Subject Headings

An author-title citation is generated for each subject heading assigned to a document reference. The maximum number of citations that may be generated is ten per reference. Any subject headings after the tenth assigned to a reference will be bypassed. Each printed line indicates a bypassed subject heading.

TAPE LIBRARY AND REPORT CONTROLS, RESTART PROCEDURES

The successful operation of a computer system depends in part on the ability to use the computer programs correctly to produce the appropriate results or products. However on a continuing basis, successful operation of the computer programs is not sufficient to guarantee success. The user must also be able to manage (i.e., use and then store or discard) the tapes and reports not for one cycle but on a continuing basis. Certain products must be saved for the completion of an individual cycle; others must be saved for several generations to provide back-up for restart, others for internal or historical purposes.

Each program module will be examined to provide the user with an understanding of the role that each report and tape file plays and the discussion that each should receive. URBANDOC used the "grandfather-father-son principle" for the maintenance and security of tape files in which three successive generations of major tape files are retained so that a current file could be recreated in case of physical damage or operator error.

Thesaurus Module

The following table illustrates the tape files produced by the Thesaurus Module along with present URBANDOC policy on their retention:

File Name	*When Created*	*When Used*	*Disposition*
Thesaurus Input File	XMAIN1	SORTX1	Scratch at end of cycle
Sorted Thesaurus Input File	SORTX1	XMAIN2	Scratch at end of next update cycle
Thesaurus File	XMAIN2	XMAIN3, XMAIN4, XMAIN6, XMAIN2	Tape Library, can scratch after third generation
Permuted Thesaurus File	XMAIN5		Tape Library
Thesaurus Supplement File	XMAIN6	XMAIN3	Tape Library

The key to this module is the Thesaurus File. With the existence of the most current Thesaurus File, any other tape file can be recreated. The Thesaurus File can be recreated using the Thesaurus File "father" and the current Sorted Thesaurus Input File. The Sorted Thesaurus Input File could of course be recreated from the card input.

The Permuted Thesaurus File and Thesaurus Supplement File play limited roles in the system. The Permuted Thesaurus File should be saved if the user intends to update it. As an alternative procedure, the Permuted Thesaurus File could periodically be recreated

(i.e., every six months or every year). The Thesaurus Supplement File need not be saved since it is usually used only once and can always be recreated.

The reports produced by the Thesaurus Module are:

Pre-list of the Thesaurus Input;
Various edit listings from the update cycle;
URBANDOC Thesaurus, either subject or geographic;
Statistical Analysis of the Thesaurus;
Permuted Thesaurus;
Thesaurus Supplement.

The Pre-list should be saved for historical purposes. The various edit listings can either be discarded or saved for historical purposes; this is left to the discretion of the user. The same holds true for the Statistical Analysis and the Thesaurus Supplement. The various editions of the Thesaurus and the Permuted Thesaurus should be saved for historical purposes. As the reports age the decision to purge the files of historical listings is left to the user.

Pre-edit Module

The following table illustrates the tape files produced by the Pre-edit Module along with present URBANDOC policy on their retention:

File Name	*When Created*	*When Used*	*Disposition*
Document Input File	E0010	E0015, E0020	Scratch at end of cycle
Sorted Document Input File	E0015	E0020	Scratch at end of cycle
Full Document Input File	E0020	E0030, E0040 P0010	Scratch at end of next cycle
Updated Document Input File	E0040	P0010	Tape Library

In the Pre-edit Module, there is no need to retain any of the tape files except as they provide additional security for the Document Master File and Publication File. The Document Input File in its fully expanded and edited format provides the additional backup for the File Maintenance Module and the Publications Module. Once the Full Document Input File has been updated with the revisions from the Editing and Validation Cycle it may be scratched. The Document Input File and Sorted Document Input File need not be saved since they can always be recreated from the card input.

The reports produced by the Pre-edit Module are:

Formatted Listing of the Document Input;

Pre-edit Listing of the Document Input;
Revisions to the Document Input.

These three listings are necessary tools for the Editing and Validation Cycle. With the completion of file maintenance, if these reports are saved, they serve only historical or audit trail purposes. The decision to purge the file of historical listings is left to the discretion of the user.

File Maintenance Module

The following table illustrates the tape files produced by the File Maintenance Module along with present URBANDOC policy on their retention:

File Name	*When Created*	*When Used*	*Disposition*
Docu-to-Tape File	MAINT1	MAINT3	Scratch at end of cycle
Look-Up Input File	MAINT1	SORTM1	Scratch at end of cycle
Sorted Look-Up Input File	SORTM1	MAINT2	Scratch at end of cycle
Bad Look-Up File	MAINT2	SORTM2	Scratch at end of cycle
Sorted Bad Look-Up File	SORTM2	MAINT3	Scratch at end of cycle
Edit File	MAINT3	MAINT4	Scratch at end of next update cycle
Print File	MAINT4	MAINT6	Scratch at end of cycle
Descriptor Input File	MAINT6	SORTD1	Scratch at end of cycle
Sorted Descriptor Input File	SORTD1	DMAIN1	Scratch at end of next update cycle
Document Master File	MAINT4	MAINT8, MAINT9, P0080, PHASE5	Tape Library, can scratch after third generation
Document Master Subset File	MAINT9	MAINT8, MAINT9, P0080, PHASE5	Tape Library
Inverted File	DMAIN1	DMAIN2, DMAIN3 SEARCH	Tape Library, can scratch after third generation

File maintenance processing occurs in three phases. At the completion of each phase, one

or two tape files summarize the various runs. These files are the Edit File, Sorted Descriptor Input File, Document Master File, and Inverted File.

The first phase creates the Edit File. No files prior to the creation of this one need be saved since they are really recreated from the Updated Document Input File. The second phase uses the Edit File to produce the Updated Document Master File and the Sorted Document Input File. The latter file is the input to the third phase which updates the Inverted File to agree with the Document Master File.

For the File Maintenance Module, the key files are the Document Master File and the Inverted File. With the existence of the most current Document Master File, the Document Master Subset File, Sorted Descriptor Input File and Inverted File may be recreated. The most current Document Master File could be recreated from the Document Master File "father" and the most current Edit File. The Document Master Subset File is usually not saved since it serves a limited purpose. The effort required to retain it outweighs that required to recreate it.

The reports produced by the File Maintenance Module are:

Edit Listing,
Non-Processed Changes;
Sequence Error Listings,
File Maintenance Listing;
Active sections of the Library Print Listing;
Statistical Analysis of the Document Master File;
Summary and Detail Listings of the Inverted File.

The various edit listings are most active during the Document Master File revision process. Once the revisions have been made, the listings need not be saved for other than historical reasons. The File Maintenance Listing and the Statistical Analysis from each cycle should be saved for audit trail and historical purposes. Sections of the Library Print Listing that have been replaced with new editions need not be saved. The decision to save older copies of the Summary Listing and Detail Listing of the Inverted File is one that may be left to the discretion of the individual user. The decision to purge the file of historical listings is also left to the user.

Search Module

The following table illustrates the tape files produced by the Search Module along with present URBANDOC policy on their retention:

File Name	*When Created*	*When Used*	*Disposition*
Table File	SEP1	SORT SEP1	Scratch at end of cycle
Sorted Table File	SORT SEP1	SEP2	Scratch at end of cycle
Table File w/Frequencies	SEP2	SORT SEP2	Scratch at end of cycle

Sorted Table File w/ Frequencies	SORT SEP2	SEP3	Scratch at end of cycle
Request File	SEP1	SEP3, PHASE6 PASS1	Tape Library
CFS Request Input File	SEP3	PHASE1	Scratch at end of cycle
CFS Request File	PHASE1	PHASE2, PHASE4 PHASE6	Tape Library
Primary Selection File	PHASE2	PHASE3	Scratch at end of cycle
Reduced Primary Selection File	PHASE3	PHASE4	Scratch at end of cycle
Inquiry File	PHASE4	PHASE5	Scratch at end of cycle
Selection File	PHASE5	PHASE6	Tape Library

The key tapes for the Search Module are the Request File, CFS Request File and the Selection File. Generally speaking it is unlikely that these tapes will be retained past the completion of the *Retrieval Report*. For specialized bibliographies or issues of the *Input Index*, these tapes should be retained for historical purposes and possible reruns.

The Search Module produces two reports:

Various error listing;
Retrieval Report.

The various error listings need not be saved past the production of the *Retrieval Report*. The *Retrieval Report* should be saved for historical purposes.

Publications Module

The following table illustrates the tape files produced by the Publications Module along with present URBANDOC policy on their retention:

File Name	*When Created*	*When Used*	*Disposition*
Publication File	P0010	P0025	Scratch at end of cycle
Content Analysis File	P0010	P0015	Scratch at end of cycle
Sorted Content Analysis File	P0015	P0050	Tape Library
Sorted Publication File	P0025	P0040, P0050 P0055	Tape Library

Subject Subset File	P0040	P0045	Scratch at end of cycle
Sorted Subject Subset File	P0045	P0050	Tape Library
Publication Index File	P0055	P0060	Tape Library
Formatted Page File	P0060	P0065	Scratch at end of cycle
Sorted Formatted Page File	P0065	P0070	Tape Library
Special Publication File	P0080	P0025	Scratch at end of cycle
Sorted Special Publication File	P0025	P0040, P0045 P0055	Tape Library

The key files to be retained past the completion of the cycle are the Sorted Content Analysis File (if created), Sorted Publication File, Sorted Subject Subset File, Publication Index File and Sorted Special Publication File. Once the printed listings and indexes have been created, except for a cumulative set of indexes, these tapes are not used again. It is recommended that these tapes be retained until their further use is doubtful. The nature of IBM 1401 COBOL makes the recreation of the tapes burdensome from a machine usage point of view.

The reports produced by the Publications Module are:

The various error listings;
Main Document Listing;
Major Subject Listing;
Other Specialized Indexes;
Corporate and Project Names Index,
Personal Author Index,
Significant Title Index,
Geographic Index,
Consultant Index, etc.

The various error listings need not be saved past the completion of the cycle. For published issues of the *Input Index*, all other listings should be retained for historical purposes. For listings and indexes produced as internal tools, the older versions may be replaced by the current editions. The decision to purge the files of historical listings is left to the user.

TIMING

Essential to systems design and operation is the ability to estimate the computer time required to operate the system. Using the data bases for the prototype files and a normal file maintenance cycle, comparative run times have been provided for each run. The basic characteristics of the files were as follows:

an average of 150 references in the prototype;
an average of 150 references in the file maintenance cycle;
an average of 3 subject headings assigned per reference;
an average of 13 descriptors assigned per reference;
an average of 10 lines of bibliographic information per reference;
an average of 25 units per reference;
an average of 15 units of keyed input per reference;

Some runs will increase arithmetically with a larger data base. Many will change by small increments. For the sorts, excerpts from the IBM SORT7 timing manual will be provided. For all procedures except searching, the time to complete a cycle will be the total of times for the individual runs. For the *Retrieval Report* Production, individual times for the SEP programs must be added to the total time for the CFS portion of the search. A timing formula has been provided for the latter.

In all instances timings are approximate and include set-up time. The computer used for the samples was the 1401 computer already described in the Introduction. The speed of the printer is 600 lines per minute, the tape drives the 7330 model. If the system is run on a configuration with faster memory, more or faster tape drives or printer, the times provided in the sample can be used as maximum times for a similar size sample.

Thesaurus Module

Note: The times for the Thesaurus Module runs are a function of the number of thesaurus changes and not of the number of references in a cycle. Any run involving the use of the full Thesaurus File will lengthen in time as the file grows in size.

Run Number	*Prototype*	*File Maintenance*	*Comments*
XMAIN0	10 mins.	10 mins.	
XMAIN1	5 mins.	5 mins.	
SORTX1	10 mins.	10 mins.	Record size of 1239 characters. See Table 7.
XMAIN2	10 mins.	10 mins.	Run time will lengthen by small increments as the Thesaurus File increases.
XMAIN3	50 mins.	50 mins.	Subject Thesaurus 20 mins. Geographic Thesaurus 30 mins. Run time directly proportional to Thesaurus File size.

XMAIN4	10 mins.	10 mins.	Run time will lengthen by small increments as the Thesaurus File increases.
XMAIN5	45 mins.	45 mins.	
XMAIN6	10 mins.	10 mins.	Run time will lengthen by small increments as the Thesaurus File increases.

Pre-Edit Module

Note: Except for the sort (E0015), all run times are directly proportional to the number of items in the input.

Run Number	*Prototype*	*File Maintenance*	*Comments*
E0010	15 mins.	50 mins.	
E0015	10 mins.	15 mins.	Record size of 84 characters. See Table 4.
E0020	55 mins.	2 hrs, 45 mins.	
E0030	15 mins.	50 mins.	
E0040	20 mins.	1 hr.	

File Maintenance Module

Note: The time for the File Maintenance Module run involving the full Document Master File will lengthen as the file grows in size.

Run Number	*Prototype*	*File Maintenance*	*Comments*
MAINT1	10 mins.	10 mins.	
SORTM1	10 mins.	15 mins.	Record size of 56 characters. See Table 3.
MAINT2	5 mins.	10 mins.	Run time will increase slightly as the Thesaurus File increases in size.
SORTM2	5 mins.	10 mins.	Record size of 59 characters. See Table 3.
MAINT3	10 mins.	20 mins.	
MAINT4	10 mins.	25 mins.	Run time will increase by small increments as the Document Master File increases.

MAINT6	25 mins.	40 mins.	
MAINT7	10 mins.	25 mins.	Run time proportional to the size of the Document Master File.
MAINT8	10 mins.	15 mins.	Run time will increase slightly as the Document Master File increases
MAINT9	5 mins.	5 mins.	Run time will increase slightly as each section of the Document Master File increases.
SORTD1	10 mins.	15 mins.	Record size of 41 characters. See Table 1.
DMAIN1	10 mins.	10 mins.	Run time will increase slightly as the Inverted File increases.
DMAIN2	10 mins.	10 mins.	This run is a function of the number of unique descriptors used, rather than of the total number of descriptors assigned.
DMAIN3	15 mins.	25 mins.	This run is directly proportional to the number of descriptors assigned.

Search Module

Note: The search run must be timed in two portions: the SEP programs as individual runs and the CFS search according to a formula. The search runs are a function of the number of questions submitted and the number of descriptors in each question. Run times are only slightly affected by the size of the Document Master File and the Inverted File, specifically SEP2, PHASE2 and PHASE5.

Run Number	*Prototype*	*File Maintenance*	*Comments*
SEP1	5 mins.	5 mins.	
SORT SEP1	5 mins.	5 mins.	Record size of 90 characters. See Table 5.
SEP2	10 mins.	10 mins.	This run will increase only slightly as the Inverted File increases.
SORT SEP2	5 mins.	5 mins.	Record size of 90 characters. See Table 5.
SEP3	10 mins.	10 mins.	

Search Run Formulas (assumes 729, Mod II, tape drives)[1]

Phase I: 15 Sec + 2 Sec (R)

Phase II: $D_F(13 + .016\ C_D)$ ms

Phase III: 2.2 RD ms for 2 descr/term

Phase IV: (100 D ± 14 RD) ms + 1 min.

Phase V: 30 B ms + 60 D ms

Phase VI: 300 lines per minute

Symbols: B – total number of records in the Bibliographic File

C_D – total number of characters in Descriptor File tape record

D – total number of search documents*

D_F – total number of tape records in the Descriptor File

F – assumed factor of reduction by "and"-ing two document no. sets

R – total number of request terms

S – total number of search descriptors

*Note: "Total number" means the number entering that particular phase. For instance, the use of 2 descriptors per request term in Phase III results in a lower total number entering Phase IV than were handled in Phase III. Phase IV and V will always have the same number of search documents.

Sample Times (in minutes)

Number of Request Terms	File Size			
	10K	25K	50K	100K
10	20-20**	30-30	42-42	74-74
25	22-22	34-34	49-48	89-86
50	37-34	57-52	76-70	141-129
100	86-75	119-103	174-150	334-287

1 International Business Machines Data Processing Division Program Information Department, *The Combined File Search System* by Donald Prentice, 18 p. (White Plains, N.Y., 1964), p. 13.

Assumptions: 1) 2 descriptors per request term are used in the Descriptor File search**

2) Average number of item postings in the Descriptor File:

Master File Size	10K	25K	50K	100K
Aver. No. of Items	50	75	100	200

**The first number assumes that the "and" operation reduces the item numbers entering Phase IV to 1/5 the average number of postings. The second number assumes a similar reduction to 1/10.

Publications Module

Note: Except for the sorts (P0015, P0045, P0055 and P0065), all run times are directly proportional to the number of items in the input.

Run Number	*Prototype*	*File Maintenance*	*Comments*
P0010	1 hr.,20 mins.	5 hrs., 50 mins.	Creation of the optional tape output does not significantly affect run time.
P0015	5 mins.	10 mins.	Record size of 38 characters. See Table 2.
P0025	10 mins.	20 mins.	Record size of 409 characters. See Table 6. This file will be mostly in sequence and therefore the times given in the table may be considered to be the maximum.
P0040	5 mins.	25 mins.	
P0045	15 mins.	45 mins.	Record size of 409 characters. See Table 6.
P0050	30 mins.	2 hrs, 30 mins.	
P0055	15 mins.	50 mins.	Record size of 409 characters. See Table 6.
P0060	50 mins.	3 hrs, 30 mins.	
P0065	5 mins.	10 mins.	Record size of 69 characters. See Table 4.
P0070	10 mins.	25 mins.	

Sort Timing Tables[2]

INPUT FILE SIZE	1401				1460							
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI	G	B	P
1000	1	1	1	1	1	1	1	1	1	39	39	5
2000	3	2	2	2	2	1	1	1	1			6
5000	8	6	5	5	6	4	4	4	3			6
10000	18	13	11	12	14	9	8	8	7			9
25000	49	36	31	32	39	26	21	22	19			10
50000	108	77	67	70	86	56	46	48	41			11
75000	161	119	104	105	129	87	72	72	61			11
100000	233	166	145	153	187	120	99	107	90			12

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 8K; Record length — 40; Control field — 10

Table 1.

INPUT FILE SIZE	1401				1460							
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI	G	B	P
1000	1	1	1	1	1	1	1	1	1	73	73	4
2000	2	2	1	1	2	1	1	1	1			5
5000	7	5	5	5	5	4	3	3	3			7
10000	16	11	10	10	12	8	7	7	6			8
25000	44	31	28	28	34	22	19	19	16			9
50000	95	69	61	62	75	49	41	42	36			10
75000	156	116	102	100	123	83	70	68	58			11
100000	207	150	132	138	164	106	89	95	81			11

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 12K; Record length — 40; Control field — 10

Table 2.

INPUT FILE SIZE	1401				1460							
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI	G	B	P
1000	2	1	1	1	2	1	1	1	1	32	16	5
2000	4	3	2	3	4	2	2	2	2			6
5000	14	9	8	8	12	7	6	6	5			8
10000	31	21	17	18	26	16	13	14	11			9
25000	85	56	47	50	72	44	35	38	30			10
50000	185	121	102	110	158	95	75	84	67			11
75000	300	189	158	171	256	146	116	128	103			12
100000	400	247	207	222	344	190	150	166	133			12

2-WAY BALANCED MERGE WITH TAPE LABEL PROCESSING
Core storage capacity — 8K; Record length — 80; Control field — 10

Table 3.

[2] International Business Machines, Data Processing Division, *Sort 7 Timing Specifications and Operating Procedures IBM 1401 and 1460 Data Processing Systems,* 26p. (White Plains, N.Y., 1964), pp 11-12, 14-15.

INPUT FILE SIZE	1401				1460					G	B	P
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI			
1000	2	1	1	1	1	1	1	1	1	36	36	5
2000	4	3	2	2	3	2	2	2	1			6
5000	13	9	7	7	11	7	5	5	4			8
10000	28	19	16	16	24	14	12	12	10			9
25000	78	52	44	45	66	40	32	33	27			10
50000	171	113	96	102	145	87	70	76	62			11
75000	270	175	149	157	236	133	107	115	94			12
100000	370	228	194	204	315	173	138	148	121			12

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 12K; Record length — 80; Control field — 10

Table 4.

INPUT FILE SIZE	1401				1460					G	B	P
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI			
1000	3	2	2	2	2	2	1	1	1	15	15	7
2000	6	4	4	4	6	3	3	3	2			8
5000	18	12	10	10	16	9	7	8	6			9
10000	40	26	21	22	34	21	16	17	14			10
25000	108	75	62	61	94	61	48	47	37			11
50000	235	147	122	132	205	116	91	101	81			12
75000	380	229	190	204	331	180	141	155	123			13
100000	507	299	248	267	442	234	183	201	159			13

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 8K; Record length — 100; Control field — 10

Table 5.

INPUT FILE SIZE	1401				1460					G	B	P
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI			
1000	12	7	6	6	11	6	5	5	4	10	5	7
2000	27	16	13	13	24	14	10	11	8			8
5000	74	44	35	36	68	38	28	30	23			9
10000	164	94	75	81	150	80	61	67	52			10
25000	487	254	201	216	446	213	160	176	133			12

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 12K; Record length — 500; Control field — 10

Table 6.

INPUT FILE SIZE	1401				1460					G	B	P
	7330	729-II	729-IV	729-V	7330	729-II	729-IV	729-V	729-VI			
1000	33	20	16	17	31	18	13	15	11	2	1	9
2000	74	45	35	38	68	39	29	32	24			10
5000	218	125	97	110	202	109	81	93	70			12
10000	472	252	195	219	437	217	161	184	137			13

2-WAY BALANCED MERGE WITHOUT TAPE LABEL PROCESSING
Core storage capacity — 8K; Record length — 1000; Control field — 10

Table 7.

LOCAL IMPLEMENTATION

The implementation of a local version of URBANDOC's computer system requires some planning and preliminary processing work by the user. Since URBANDOC's total system contains the two major goals of publications and searching, the user must first establish toward which of these goals he will work first. It is URBANDOC's recommendations that a local system start with the publications function. This approach has been described in the *General Manual* as the Local Systems Module. As such, it involves the Pre-edit Module and the Publications Module. A second phase of the local effort could later expand the system to include the search function. These recommendations are based on the more complex nature of the search function both from an intellectual and data processing viewpoint.

Once the direction of the local installation has been determined, the user must establish a program library for the source and object programs to be used. If searching is part of the goal, a systems tape containing the monitor and the search object programs must be created. Contrary to the indications in the CFS documentation, URBANDOC has not found it necessary, or even desirable, to operate the other program modules in a monitored environment for a four tape system. This decision is based on more cumbersome operations resulting from a continual mounting and dismounting of the systems tape when all four drives are required by the processor program.

BIBLIOGRAPHY

Application Programs

Engineering Index, Inc. *1968 User Participation Program Level II (Tape Service) System Documentation and User's Manual for File Search.* New York (Sept. 1967) 1 v.

International Business Machines, Data Processing Division. *1401 Information Storage and Retrieval System (The Combined File Search System)* by Donald Prentice, Gary deGraw, Alice Smith and I. Albert Warheit. White Plains, N.Y. (May 1, 1966) 1 v. 1401 General Program Library #10.3047 Rev. III

_____.*The IBM 1401 Information Storage and Retrieval Program for Engineering Parts and Drawings.* White Plains, N.Y. (n.d.) 22p. Application Manual #E20-0132

_____.*Quick KWIC Index,* by J. M. Bain. White Plains, N.Y. (n.d.) 14p. 1401 General Program Library #10.3.035.

Programming Systems

International Business Machines, Data Processing Division. *Autocoder (on Tape) Language Specifications and Operating Procedures IBM 1401 and 1460.* White Plains, N.Y. (1961) 60p. Form #C24-3319.

_____.*Autocoder (on Disk) Language Specifications IBM 1401, 1440, and 1460.* White Plains, N.Y. (1964) 57p. Form #C24-3258.

_____*Autocoder (on Disk) for IBM 1401/1460 Specifications.* White Plains, N.Y. (1962) 49p. Form #C24-1486.

_____*Autocoder (on Disk) Program Specifications and Operating Procedures.* White Plains, N.Y. (n.d.) 65p. Form #C24-3259.

_____.*Common Business Oriented Language (COBOL) General Information.* White Plains, N.Y. (1960) 163p. Form #F28-8053.

_____.*COBOL (on Disk) Specifications IBM 1401, 1440, and 1460.* White Plains, N.Y. (1964) 46p. Form #C24-3235.

_____.*COBOL (on Disk) Program Specifications and Operating Procedures IBM 1401, 1440, and 1460.* White Plains, N.Y. (1964) 61p. Form #C24-3242.

_____.*Input/Output Control System (on Disk) Specifications IBM 1401 and 1460.* White Plains, N.Y. (1962) 69p. Form #C24-1489.

_____.*Sort 7 Specifications and Operating Procedures IBM 1401 and 1460.* White Plains, N.Y. (1963) 55p. Form #C24-3317.

_____.*Sort 7 Timing Specifications and Operating Procedures IBM 1401 and 1460 Data Processing Systems.* White Plains, N.Y. (1963) 26p. Form #C24-1456.